International Journal on Advances in Life Sciences

2025 vol. 17 nr. 1&2

The *International Journal on Advances in Life Sciences* is published by IARIA.
ISSN: 1942-2660
journals site: http://www.iariajournals.org
contact: petre@iaria.org

Sponsored by IARIA
www.iaria.org

International Journal on Advances in Life Sciences
Volume 17, Number 1 & 2, 2025

International Journal on Advances in Life Sciences
Volume 17, Numbers 1 & 2, 2025

CONTENTS

Integrating and Analysing Occupational Health Data Using a Multi-Ontology Approach

Cassandra BARBEY
Department of Pollutant Metrology, INRS
Univ Rennes, Inserm, EHESP, Irset- UMR_S 1085
Vandœuvre-lès-Nancy, France
Email: cassandra.barbey@inrs.fr

Nathalie BONVALLOT
Univ Rennes, Inserm, EHESP, Irset - UMR_S 1085
Rennes, France
Email: nathalie.bonvallot@ehesp.fr

Malika SMAÏL-TABBONE
Université de Lorraine, CNRS, LORIA, UMR 7503,
Vandœuvre-lès-Nancy, France
Email: malika.smail@loria.fr

Frédéric CLERC
Department of Pollutant Metrology, INRS
Vandœuvre-lès-Nancy, France
Email: frederic.clerc@inrs.fr

***Abstract* - A variety of occupational data are collected by health organisations to investigate workplace exposures encountered by workers in their occupational activities and the potential health effects that may arise. These datasets have diverse characteristics and are not inherently designed to interoperate. However, they contain complementary information, which, when analysed collectively, can provide a broader perspective on high-risk occupational scenarios and inform targeted prevention strategies. The objective of this study is to develop a methodology to integrate and analyse heterogeneous French data. For this, ten French occupational databases, provided by six French institutes were used. An Ontology-Based Data Integration approach was employed, involving the mapping of data sources to a domain-specific ontology, namely the Adapted Occupational Exposure Ontology. Four additional ontologies were utilised: the Occupational Exposure Thesaurus, which categorises occupational exposures and hazards; the International Classification of Diseases, which classifies health disorders and diseases; the French Nomenclature of Activities, which identifies activity sectors in France; and the Professions and Socio-professional Categories, which defines occupational classifications. Data integration is primarily achieved through the concept of the "occupational group", defined as a group of individuals sharing the same sex, occupation, and activity sector. Two case studies derived from the integrated dataset are presented: (1) a quantitative analysis identifying occupational groups at highest risk and most affected by diseases; and (2) a qualitative analysis evaluating the consistency of exposure and disease-related information. The construction sector was selected for these case studies due to its significance in occupational health research and the availability of substantial, relevant data. This methodological approach structures all the data and enables various analysis methods to be designed and implemented, making it possible to envisage targeted responses to current and emerging occupational health problems using specialised tools and queries.**

Keywords - ontologies; data integration; heterogeneous data; occupational health; occupational exposures; data analytics.

I. INTRODUCTION

This article constitutes an extended version of the international conference paper entitled "Ontology-Based Integration of Occupational Health Data: Method and Case Studies" [1], which was presented at the Thirteenth International Conference on Data Analytics in October 2024. The indicator calculation method has been updated, and the use of case studies has been expanded to the occupation and sector of activity.

Workers are routinely exposed to various occupational hazards that may affect their safety and health, potentially leading to occupational accidents or diseases. These risks result from occupational exposures, which can be present in any work situation linked to the specific tasks performed in a given occupation. Occupational exposures are categorised into five main groups: chemical, physical, biological, organisational, and psychosocial. Chemical exposures encompass all chemical products or substances encountered in occupational settings, regardless of their form, such as paints, cleaning products, and dusts. Physical exposures refer to all exposures that may have a physical effect on the worker (*i.e.*, temperature, vibrations, or noise). Biological exposures involve contact with living organisms encountered in occupations, including microorganisms, animals, or plants. Organisational exposures relate to structural aspects of the workplace (*i.e.*, hours, workload, or lack of resources). Psychosocial exposures refer to workplace dynamics that impact psychological and social well-being (*i.e.*, stress, hierarchical relations, or bullying). The interactions among these various exposures can have complex effects on workers' health, potentially reducing the efficacy of risk mitigation measures often designed for individual exposures. The implementation of relevant preventive actions requires an understanding of these interactions and their effects for effective occupational risk management. However, this remains an understudied area despite ongoing research efforts [2].

In France, several health institutions collect data on occupational exposure and workers' health. Each database is designed for a specific objective, such as characterising work environments or monitoring occupational disease development. Consequently, each dataset possesses its own characteristics, including its collection method and target population. This diversity of characteristics enables the collection of a substantial amount of information; however, it also hinders the ability to share this information, as the databases were not designed to interact with one another. While some databases provide a representative overview of the French workforce, others have a more limited scope. For instance, the Sumer database, derived from cross-sectional surveys on workplace exposures and worker perceptions, provides a nationally representative dataset. By contrast, the Scola database, which compiles exposure measurements conducted under regulatory chemical substance monitoring, has a different scope and unit of measurement, focusing on companies subject to regulatory controls. Despite their differences, both databases contribute valuable insights into chemical exposure in occupational settings.

The French DataPOST project (analysis of multiple exposure and workers' health) aims to provide a general joint use of French occupational health data. It aims to develop a methodology for extracting knowledge on occupational exposure and health effects using an ontological approach, by integrating data from ten occupational health databases. To achieve this objective, we propose to rely on the Ontology-Based Data Integration (OBDI) approach [6] to qualify and quantify occupational exposures and associated health effects. The statistical unit for analysis is the "occupational group", which constitutes a group of individuals of the same sex sharing the same occupation and working in the same activity sector. Such occupational groups are defined across all databases, facilitating data integration and exploitation. The integrated dataset is then utilised for multiple analyses, two of which are presented (Sections V and VI):

- A **quantitative analysis**, which quantifies various exposures and diseases for each occupational group by generating relevant indicators;
- A **qualitative analysis**, which assesses data consistency across databases for each type of exposure and disease.

The remainder of the article is structured as follows: Section II presents the current status of research on the joint use of several data sources. Section III provides an overview of OBDI and its main components (ontologies, data schemas, and mappings) used in this study. Section IV details the general data representation framework. Sections V and VI present the case studies, describing the methodologies and results. Finally, Section VII concludes with a discussion of findings and potential future research directions.

Specific Author Contributions:

- Cassandra BARBEY: Writing – original draft, Methodology, Investigation, Formal analysis;
- Malika SMAIL-TABBONE: Writing – review & editing, Methodology, Investigation, Formal analysis, Validation;
- Nathalie BONVALLOT: Writing – review & editing, Validation, Supervision;
- Frédéric CLERC: Writing – review & editing, Methodology, Validation, Supervision.

II. Related Work

While the datasets under consideration are diverse in their structure and scope, their combined analysis holds the potential to provide a more comprehensive view of work-related risks and their impacts. To enable the joint use of these varied data sources, multiple analytical approaches have been developed, although their specificity limits their reproducibility. Some methodologies have been applied in case studies, such as those conducted by L. Rollin *et al.* [3], [4]. The first study compares data from multiple occupational health surveillance and monitoring systems using homecare workers as an example [3]. The second study examines musculoskeletal disorders in male professional drivers and their associated risk factors, leveraging multiple occupational health databases [4]. The use of the databases was based on the objective, the population, and the target disease. The joint analysis therefore represents a specific subset of the active working population. Other methodologies are being developed within broader projects, within the European context, such as the Datamining project, which aims to promote the monitoring of populations of workers at risk, by centralising data, in order to improve occupational sensitisation and prevention measures (https://data.risquesautravail.be/fr). This initiative integrates administrative records and survey data collected at multiple levels (European, federal, regional) to ensure that Belgian populations are represented. A related approach was employed by Dandan *et al.* [5]; however, their work was not conducted within occupational health, instead studying the monitoring of elderly health, where ontologies were used to integrate data from sensors, surveys, and medical records.

III. Ontology-Based Data Integration (OBDI) Approach

The OBDI system aims to achieve semantic integration of heterogeneous data sets by leveraging an existing descriptive ontology. This system establishes relationships between theoretical concepts in a specific field and real-world data using formal definitions crafted by experts in the domain. The OBDI framework consists of three key components: the domain ontology, which encapsulates the descriptions, definitions, and concepts relevant to the domain; the data sources, which represent diverse storage systems containing real-world data; and the mapping, which provides the structure for reconciling raw data with theoretical concepts in the ontology through formal definitions (Figure 1). The OBDI system enhances the understanding and utilisation of data, offering reproducibility and adaptability across various domains through a wide range of semantic integration approaches [7], [8]. An illustrative example is provided in R. Dandan's thesis [9], where the aim was to design health profiles for the elderly to improve the associations of health recommendations using a knowledge-based integration approach (DIKG2). In this context, an ontology of activities for older individuals was developed, incorporating preventive strategies and recommendations for nutrition and physical activity. Data were gathered from patients' medical records, connected measurement devices, and online questionnaires. The integration of these data was facilitated through vocabulary correspondence (*i.e.*, What pathology do you have? → Diabetes = Disease class: Diabetes) as well as a link between two entities (*i.e.*, John has diabetes → Diabetes recommendation).

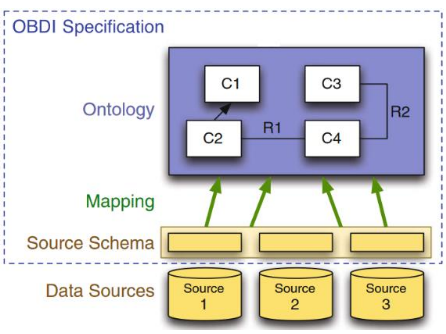


Figure 1. Ontology-based data integration adapted from Calvanese *et al.* (2017) [6].

The OBDI approach serves as the methodological foundation for the present study. However, this study's context is more complex, necessitating the use of multiple ontologies to integrate the available data and their interrelationships. For this purpose, we decided to start with the Exposure Ontology (ExO) proposed by Mattingly *et al.* [10].

A. *Occupational health ontologies*

The ExO ontology comprises four core exposure concepts: receptor, stressor, event, and outcome. Each of these concepts is further detailed through child terms and attributes. The "receptor" refers to an individual worker or a population of workers who may be exposed to a stressor. The "stressor" represents exposure to an agent, activity, or event that has the potential to impact the receptor. Stressors are categorised into chemical products, microbiological agents, physical elements, postural constraints, organisational or psychosocial factors, or combinations of these. The interaction between the stressor and the receptor is termed an "exposure event", which may result in a health "outcome", such as disease (see Figure 2, left-hand part).

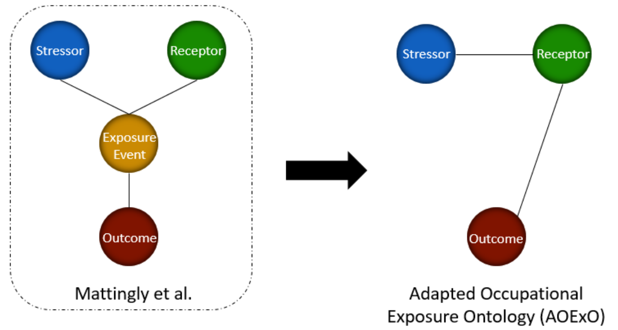


Figure 2. Main concepts of the exposure science ontology from Mattingly *et al.* (2012) [10] and their adaptation to our context.

In this study, the receptor corresponds to the occupational group, and the stressor represents occupational exposure. Unlike ExO, the "exposure event" concept is not explicitly defined and, therefore, is not used in our context. However, the available data allow the direct association of occupational groups with health outcomes. This leads to the Adapted Occupational Exposure Ontology (AOExO) (see Figure 2, right-hand part).

Four ontologies from the field of occupational health have been linked to AOExO concepts to create a more comprehensive occupational health database:

- The Occupational Exposure Thesaurus (TEP) [11] is a reference system designed in 2014 by the French agency for health safety to uniformly collect data on occupational exposures. The TEP is structured across eight hierarchical levels, encompassing approximately 8,300 exposure concepts, including "industrial product or process", "quality of the work space", "equipment tools machines and work machinery", "chemical agents", "biological agents", "rocks and other mineral substances", "physical agents", "biomechanical factors", and "organisational interpersonal and ethical factors". The system classifies exposures from general to specific concepts (*i.e.*, exposure to physical agents in general → exposure to biomechanical factor → exposure to work with strength → exposure to handling loads and handling people → exposure to carrying a load (horizontal movement) → exposure to carrying loads above the level of the shoulder).
- The International Classification of Diseases, tenth revision (ICD-10) (https://www.who.int), designed and maintained by the World Health Organisation, is an international compilation on the causes and consequences of human disease. ICD-10 provides a common health language by using approximately 150,000 codified clinical terms [12]. The classification is divided into 22 chapters, which are further subdivided into blocks of three-character categories and, in some cases, four-character subcategories. These diseases are categorised from broad concepts to more specific ones such as "Diseases of the musculoskeletal system and connective tissue" (Chapter XIII),

"Dorsopathies" (M40-M54), "Other dorsopathies" (M50-M54), "Dorsalgia" (M54), and "Sciatica" (M54.3).

- The statistical classification of activity sectors in the European Community is used for the organisation of information pertaining to economic and social activities. In the context of this study, the French version of this classification, referred to as NAF, is employed to define occupational groups (https://www.insee.fr). NAF is organised into five nested levels, such as "Construction" (Section F), "Specialised construction activities" (Division 43), "Demolition and site preparation" (Group 43.1), and "Demolition" (Class 43.11 and Sub-class 43.11Z).
- The Professions and Socio-professional Categories (PCS) (https://www.nomenclature-pcs.fr), a statistical classification developed by the French National Institute of Statistics and Economic Studies, groups occupations by social background. This ontology is used to define occupational groups. PCS is organised into four hierarchical levels of job designations, such as "Manual workers" (Aggregate category 6), "Skilled manual workers" (Category 61), "Skilled artisanal manual workers" (Detailed category 63), and "Skilled bricklayers" (Occupation 632a). Additionally, an intermediate level not officially included in the nomenclature is available, such as "Skilled craftsmen in building" (632). The latter is used to create occupational groups as it provides an optimal balance between a sufficiently small number of occupations and an appropriate level of detail.

The development of the AOExO ontology has been guided by a set of competency questions arising directly from the problems encountered in occupational health. These questions guided the choice of relevant vocabularies and ontologies to be integrated. In particular, they cover occupational exposure, work activities (occupation and sector) and health effects. These questions made it possible to precisely define the scope of the ontology, by identifying the essential concepts to be represented as well as the levels of granularity required to respond to the issues. Based on these questions, the relationships between the concepts were constructed in such a way as to reflect the complexity of real work situations, taking into account the links between occupational groups (modelled on the basis of NAF and ISCO nomenclatures), exposures (defined on the basis of TEP) and health effects (classified by ICD-10). They have also made it possible to test and validate the ontology's ability to respond to a wide range of competency questions, including the following main ones:

- "What are the differences in occupational diseases between men and women?" (Using ICD-10 ontology)
- "What are the exposures encountered by bricklayers in the workplace?" (Using TEP and ISCO ontologies)
- "Are painters in the specialised construction sector exposed to benzene?" (Using ICSO, NAF and TEP ontologies).

B. Data sources and their schema

A total of ten data sources were used in this study. Six of them are focused on occupational exposures: SUMER [3], C2P [13], COLCHIC and SCOLA [14], COLPHY [15] and MatGene [16]; three are focused on occupational diseases: AT-MP [17], MCP [3] and RNV3P [3]; and one provides both data: Evrest [3]. A comprehensive description of all data sources, collection methods, statistical units, and institutions is provided in Table I.

TABLE I. TABLE DESCRIBING THE DIFFERENT DATA SOURCES AND THE INFORMATION THEY CONTAIN

Data source	Institution	Collection method	Original statistical unit
Sumer	Dares	National surveys	Worker
C2P	CNAM	Regulatory declarations	Worker
Colchic / Scola	INRS	Sampling and analysis of workplace air by specialised chemistry laboratories	Measurement
Colphy	INRS	Historical measurements and sampling	Measurement
MatGene	SPF	Historical and census information	Occupational group
AT-MP	CNAM	Medical consultation	Worker
Evrest	Gis Evrest	Systematic occupational health interviews	Worker
MCP	SPF	Compulsory professional medical consultation	Worker
RNV3P	Anses	Medical consultation with a specialist of CCPPE	Health problem
Data source	**Target population**	**Content**	**Example**
Sumer	All French workers	340 columns representing exposures to which workers are exposed to.	Exposure to lead [yes ; no]
C2P	All workers employed by private companies	10 columns representing exposures to which employers declared the worker are exposed to.	Exposure to repetitive movements [yes ; no]
Colchic / Scola	All French companies with exposure measurements / All French companies with regulatory controls	460 columns representing the measurement of the intensity of the concentration of 230 substances in the air with regards to the regulatory limit value.	Lead Intensity [moderate ; high ; very high]
Colphy	French companies with exposure measurements in national campaign or volunteering	Four columns representing the measurement of the intensity of the emissivity of two physical exposures with regards to the regulatory limit value.	Whole body vibration Intensity [moderate ; high ; very high]
MatGene	All French workers regardless of their status or employer (salaried or self-employed, private or public)	Four columns representing the nature and intensity of exposure according to occupation.	Night work [yes ; no]
AT-MP	All French workers affiliated to the general health care system	86 columns representing occupational recognised diseases.	Spondylopathies [yes ; no]
Evrest	All employees born in October over 18 years old with at least two months of seniority	45 columns representing the percentage of workers concerns by exposure. 15 columns representing clinical signs. 15 columns representing first treatment.	Noise [yes ; no] Treatment for hearing problems [yes ; no]
MCP	All employees reported by a physician as sick due to occupational exposure	720 columns representing exposures associated with reported occupational diseases and 56 columns representing these work-related diseases.	[Allergic contact dermatitis] [Chemical agents]
RNV3P	All registered people (craftsmen, apprentices, students, disabled, self-employed, unemployed, retired…)	129 columns representing the occupational diseases identified and 908 columns representing the exposures probably linked to these diseases.	[Scoliosis] [Heavy loads ; Awkward postures]

Abbreviations: Anses: French Agency for Food, Environmental, and Occupational Health & Safety; CNAM: National Fund for Health

Insurance; Dares: Directorate of Research, Economic Studies, and Statistics; Gis Evrest: Scientific Interest Group for a monitoring system; INRS: The French National Research and Safety Institute for the Prevention of Occupational accidents and Diseases; SPF: Santé publique France.

These databases are confidential, and access to and use of them is restricted by agreement with the various health organisations that supply them.

C. *Mapping between data schemas and ontologies*

The mapping between data schemas and ontologies is performed in three stages (Figure 3):

- Defining Occupational Groups: Combinations of activity sector (NAF division), occupation (PCS intermediate level), and sex are used to create unique occupational groups, such as "Skilled craftsmen in building (PCS 632, sex male) in specialised construction activities (NAF 43)".
- Standardising Health Outcomes: Variables from the AT-MP, MCP, and RNV3P data sources are mapped to disease codes in the ICD-10 classification, grouping diseases into blocks such as "Other dorsopathies" (M50-M54).
- Identifying Occupational Exposures: Exposure variables are mapped to the TEP by comparing descriptions in the data source collection protocols with standardised descriptions of exposures in the TEP. Expert working groups were convened to validate these mappings. For example, the variable exposure "manual handling of heavy loads" in C2P, which refers to any activity necessitating the utilisation of human force (lifting, lowering, transporting object), is mapped to the TEP category "handling loads and handling people". Similar mappings were performed for all data sources (SUMER, MatGene, C2P, Evrest, COLCHIC/SCOLA, COLPHY, MCP, RNV3P).

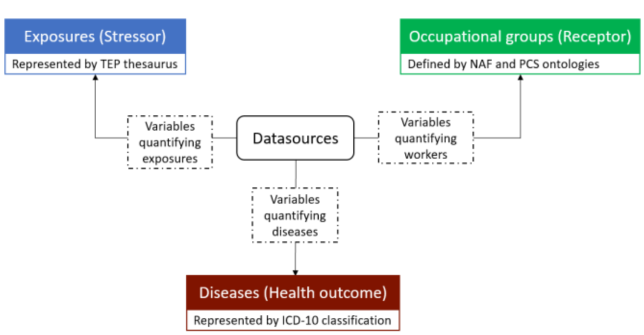


Figure 3. General representation of data and ontologies used, linked to the AOExO ontology.

IV. General representation of the health occupational data using ODBI and adapted OExO

The proposed methodology was applied to construct a total of 11,331 occupational groups, identify 88 occupational exposures, and document 191 occupational diseases. The resulting integrated dataset can be represented as a complex database comprising over 3 million exposure-disease pairings across various occupational groups, as illustrated in Figure 4. Due to the inherent limitations of the data sources, the database inevitably contains missing data. The extent of occupational risk exposure and disease development among workers is influenced by their respective occupations and the specific tasks they perform.

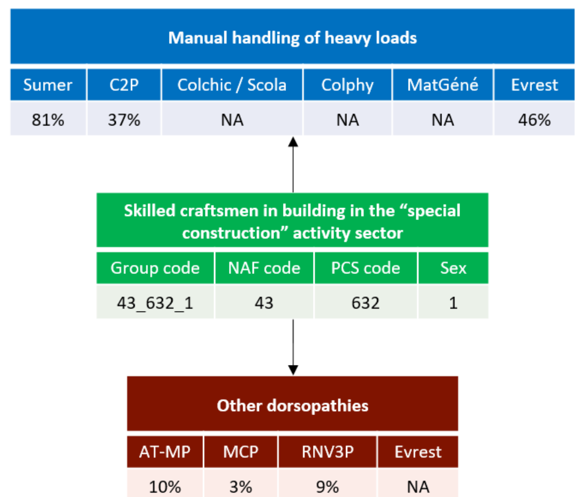


Figure 4. Example of pairing of exposure (blue – stressor) and disease (red – health outcome) for an occupational group (green – receptor).

The case studies discussed in the next two sections demonstrate practical applications of the integrated data. Each case study is defined by distinct objectives and methodological approaches including appropriate indicators. The methodologies developed can be applied to individual occupational groups or extended across multiple occupations and activity sectors.

The development of methodologies and analyses was facilitated by RStudio, an integrated development environment (IDE), using version 4.3.2 (https://docs.posit.co/ide/user/).

V. Case study one: quantitative analysis

The integration of multiple data sources can be used to facilitate the visualisation of occupational exposures and diseases of greatest concern for each occupational group. By bringing different sources of data together, it becomes possible to identify significant risks to occupational health, irrespective of the perspective from which the work situation is analysed. The availability of this information enhances the targeted implementation of safety measures aimed at mitigating exposures associated with specific work environments and reducing the risk of occupational diseases. To this end, an indicator has been constructed for each exposure pair and disease pair to each occupational group.

A. Indicator construction method for simple exposure or disease

Due to the heterogeneity of the data, the recorded values for each exposure and disease vary in scale. For example, the Sumer dataset is representative of French workers, whereas the Scola dataset originates from companies that are subject to regulatory controls (see Table 1). To standardise the data for indicator construction, values from each original source were transformed into non-parametric values and discretised on a scale from 1 to 10, where 1 represents “very low” and 10 denotes “very high”. The discretised values were then averaged based on the type of information they represent, *i.e.*, exposure or disease. The exposure indicator was calculated as the mean of the exposure intensity values across all data sources, while the disease indicator was derived as the mean of the disease intensity values from all sources.

As an illustrative example, skilled craftsmen in the “special construction” activity sector (43_632_1) may be exposed to manual handling of heavy loads, with some workers experiencing dorsopathies. For this occupational group, exposure data were available from the Sumer, C2P, and Evrest sources (Table II), while disease data were obtained from the AT-MP, MCP, and RNV3P sources. Each original value was discretised according to the data available for manual handling of heavy loads and dorsopathies. The exposure indicator was determined by averaging the values from exposure data sources, yielding a score of 7.3, whereas the disease indicator, calculated as the mean of the disease data sources, resulted in a score of 5.5.

TABLE II. SUMMARY TABLE OF DATA ON MANUAL HANDLING OF HEAVY LOADS AND DORSOPATHIES FOR THE OCCUPATION GROUP “43_632_1”

	Data source	***Original data source value (%)***	***Discretised value***	***Indicator***
Manual handling of heavy loads	Sumer	81	8	Exposure indicator: 7.3
	MatGene	/	/	
	C2P	37	4	
	Evrest	46	10	
	Colchic /Scola	/	/	
	Colphy	/	/	
Other dorsopathies	MCP	3	6	Disease indicator: 5.5
	RNV3P	9	6	
	AT-MP	10	5	

The presence of missing data across different data sources is interpreted as a lack of available information rather than an absence of exposure or disease. The Sumer dataset is the sole source that distinguishes between non-exposed workers and non-respondents. When constructing indicators, missing data are disregarded, but the information on non-exposed workers provided by Sumer is taken into account.

The constructed indicators are then represented in the form of a heatmap, where indicator values are visualised using a colour gradient. Cool colours, such as blue and green, represent lower indicator values, while warm colours, such as orange and red, indicate higher values Each heatmap focuses on a specific category of information—exposure or disease—forming the y-axis, while the x-axis represents occupational groups, activity sectors, or occupations. Three types of heatmap are presented in this case study: (1) a heatmap of the most exposed occupational groups in the construction sector, (2) a heatmap of the most exposed activity sectors, and (3) a heatmap of the most exposed occupations. In the latter two heatmaps, the indicators represent the average values for all occupational groups within the same activity sector or occupation. For example, in the case of activity sector 43, labelled “specialised construction activities”, indicators were calculated as the mean values of all occupations within this sector, including directors, secretaries, and manual workers.

B. Visualisation of exposure and disease indicators in the construction sector

As illustrated in Figure 5, the left-hand panel presents the ten most exposed occupational groups within the construction sector—a critical case study in France—while the right-hand panel displays the disease indicators for these same groups. The x-axis represents various occupational groups arranged in ascending order according to the number of exposures with those experiencing fewer exposures positioned on the left and those with greater exposure levels on the right. The y-axis illustrates the exposures or diseases, ranked according to the cumulative sum of their respective indicators. Exposures or diseases affecting a significant number of occupational groups appear towards the upper portion of the y-axis, while those impacting fewer groups are positioned lower.

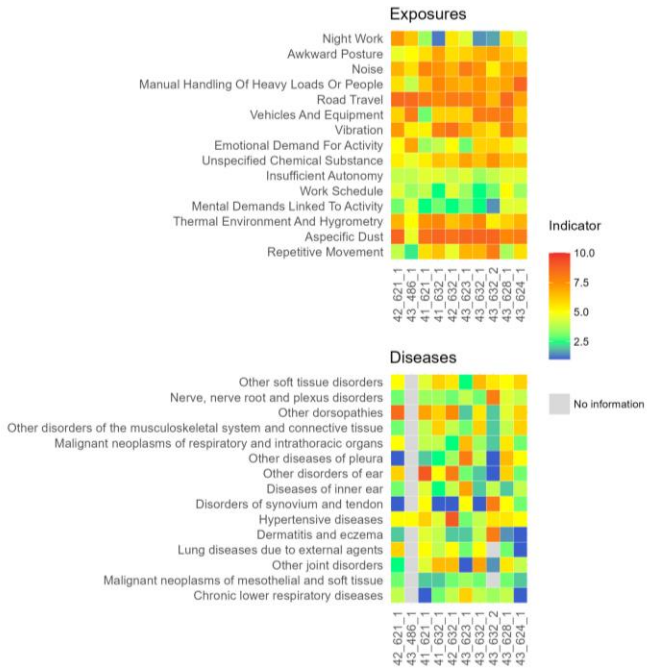


Figure 5. Heatmap of indicators by occupational group, exposure and disease: focus on 10 most exposed occupational groups and 15 exposures and diseases.

For instance, skilled craftsmen in building in the "special construction" activity sector (43_632_1 see above) were identified as the fourth most exposed and the most affected by occupational diseases, primarily due to the presence of numerous high-value indicators.

The most exposed occupational groups are particularly vulnerable to physical exposures such as noise, manual handling of heavy loads, road travel, vehicle use, and vibration. They also face exposure to certain chemical hazards, including non-specific dust, asbestos, silica, and construction products, commonly encountered in the construction industry. Furthermore, some groups are significantly affected by psychosocial and organisational risks, such as high emotional demands, limited autonomy, and mental strain. Musculoskeletal disorders (*e.g.*, soft tissue disorders, nerve root disorders), and hearing impairments (*e.g.*, inner ear diseases), are prevalent in this sector.

Although the present study does not quantify the direct correlation between exposures and diseases, it provides a qualitative assessment that underscores the need to enhance preventive measures to reduce worker exposure. This prompts reflection on the development of safety interventions tailored to work environments characterised by multiple high-risk exposures. Several ongoing studies seek to propose more effective preventive solutions [18], [19], [20].

C. *Visualisation of exposure and disease indicators across all sectors*

Figures 6 and 7 helps to identify the most exposed activity sectors alongside the most concerning exposures across all sectors. The x-axis displays all French activity sectors, labelled with letters below the heatmap, along with their corresponding divisions, represented by numbers. For example, the construction sector (labelled "F") includes divisions 41, 42, and 43. The y-axis represents the exposures identified through the data integration method, structured according to the hierarchical TEP classification, as indicated to the left of the heatmap. A detailed breakdown of the activity sector divisions and exposures can be found in Annexes A and B.

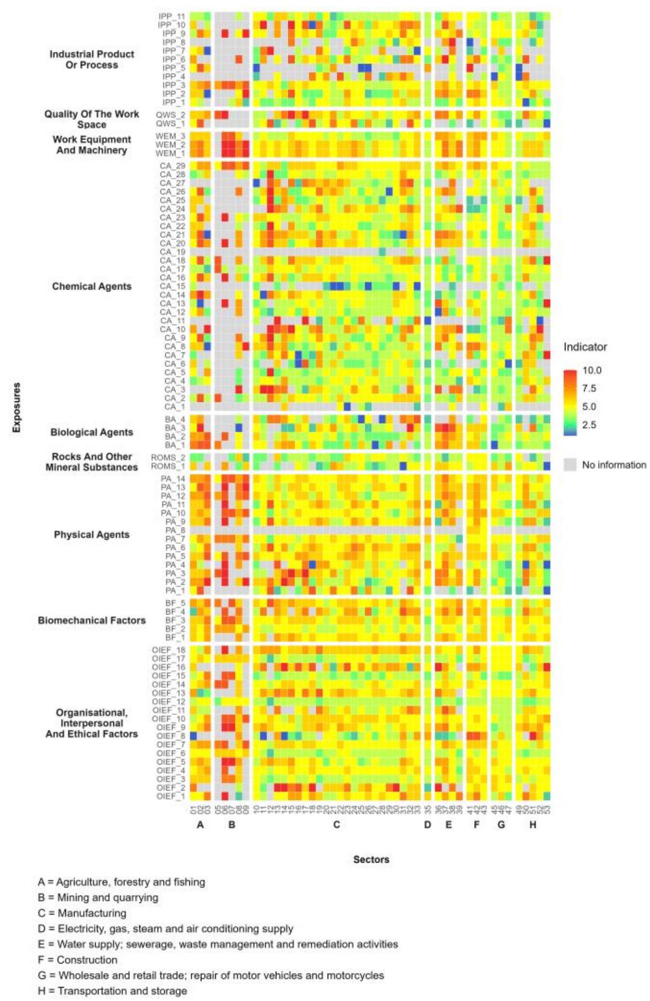


Figure 6. Heatmap of averaged indicators of all occupational groups for activity sectors A to H and by exposure.

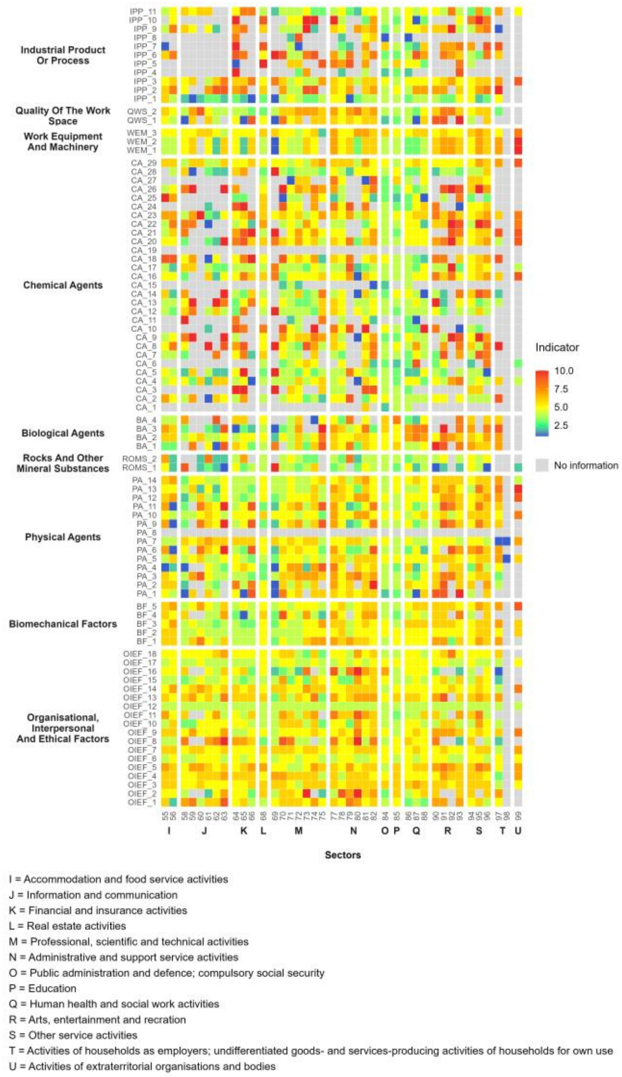


Figure 7. Heatmap of averaged indicators of all occupational groups for activity sectors I to U and by exposure.

The heatmap also reveals gaps in data availability, particularly in extractive industries (B), household employer activities (T), and extra-territorial activities (U), which include work conducted in diplomatic organisations such as embassies and consulates. Furthermore, it highlights sector-specific exposures, demonstrating the necessity for tailored preventive strategies.

For instance, workers in the agriculture sector (A) demonstrated elevated indicators of exposure to industrial materials and products (IPP_3), exposure to work equipment and machinery and some chemical agents (CA_10, CA_14, or CA_20), as well as exposure to biological agents, a preponderance of physical and biomechanical exposures, and several psychosocial-organisational factors. These observations are consistent with the findings of previous research that explored the identification of exposure profiles [21] and identified that agricultural workers are mainly exposed to organisational, chemical, physical, and biological constraints.

Another example is the sanitation and waste management sector (E), which presents high indicators for exposure to industrial products and processes, work equipment and machinery, some chemical agents (CA_3, CA_10, CA_20, and CA_21), biological agents, in particular microbiological agents (BA_3), as well as a majority of physical and biomechanical exposures, and some psychosocial-organisational factors. The results of this example are also consistent with the work of C. Fourneau *et al.* [21], who identify sanitation workers as being largely exposed to organisational, chemical, physical, and biological constraints.

The heatmap further elucidates the sectors most vulnerable to such hazards, underscoring the necessity for preventive measures tailored to the specific exposures encountered.

D. Visualisation of exposure and disease indicators across all occupations

The third heatmap of indicators (divided into Figures 8 and 9) helps to identify the occupations with the highest levels of exposure, as well as the exposures of greatest concern across all occupational categories. It highlights the disparity in available information between different occupations and their specificity. The x-axis represents all existing occupations (numbered below the heatmap) along with their corresponding intermediate levels (numbered below the heatmap). For instance, farm operators (number 1) are associated with intermediate levels 111, 122, and 131. The y-axis follows the same structure as the previous heatmap, depicting exposures identified through the data integration method, grouped according to the eight hierarchical levels of TEP. A detailed breakdown of the intermediate levels of occupations can be found in Annex C.

The figure reveals discrepancies in the availability of information across various occupations, particularly for farm operators (1) and craftsmen, shopkeepers, and company managers (2), which are associated with very few indicators. Each occupation is characterised by specific exposures linked to the tasks performed by workers, as illustrated by the heatmap.

For example, medium-sized farmers (122) exhibit elevated indicators for exposure to work equipment and machinery, as well as exposure to animals (BA_1) and biological agents (BA_2). Additionally, they face exposure to a range of physical, biomechanical, and psychosocial-organisational agents. These findings align with existing literature, which identifies farming as a high-risk occupation involving significant physical effort, frequent use of heavy machinery, and exposure to chemical substances, all of which are inherent to agricultural activities such as cultivation and livestock breeding [22], [23].

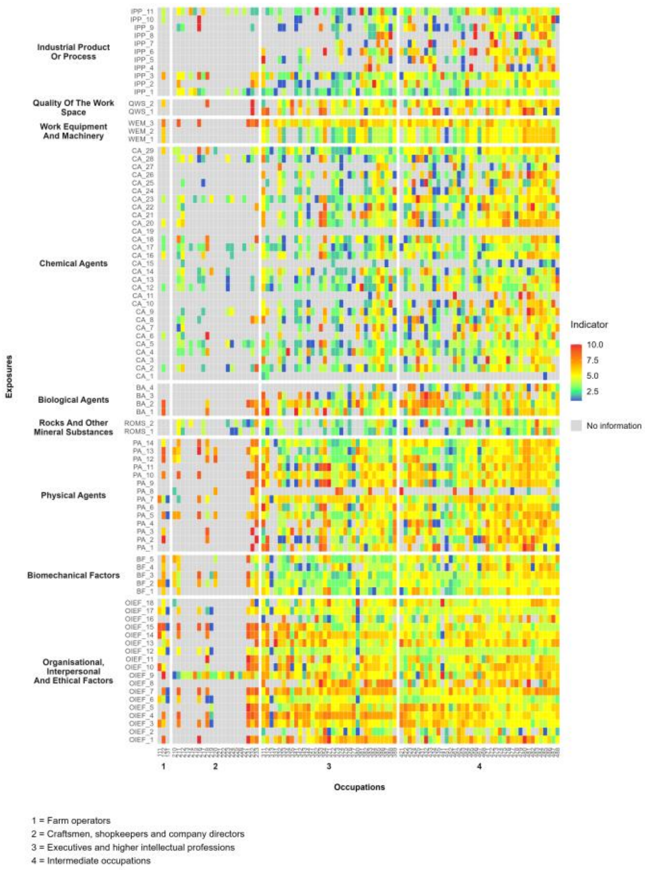


Figure 8. Heatmap of averaged indicators of all occupational groups for occupations 1 to 4 and by exposure.

Another notable example is that of police officers and prison wardens (531), who demonstrate high levels of exposure to biological agents (BA_2), as well as certain physical agents such as road travel (PA_10) and thermal environments (PA_12). Moreover, they are frequently exposed to psychosocial-organisational stressors, including external violence (OEIF_5), night work (OEIF_9), and workstation-related factors (OEIF_18). These observations are corroborated by existing research, which highlights that police officers and prison wardens routinely encounter challenging work environments. Their duties, which involve surveillance and maintaining authority, often expose them to various psychosocial stressors, including verbal and physical aggression, as well as organisational challenges such as atypical working hours and night shifts. Furthermore, they are at risk of exposure to biological substances such as blood [24], [25].

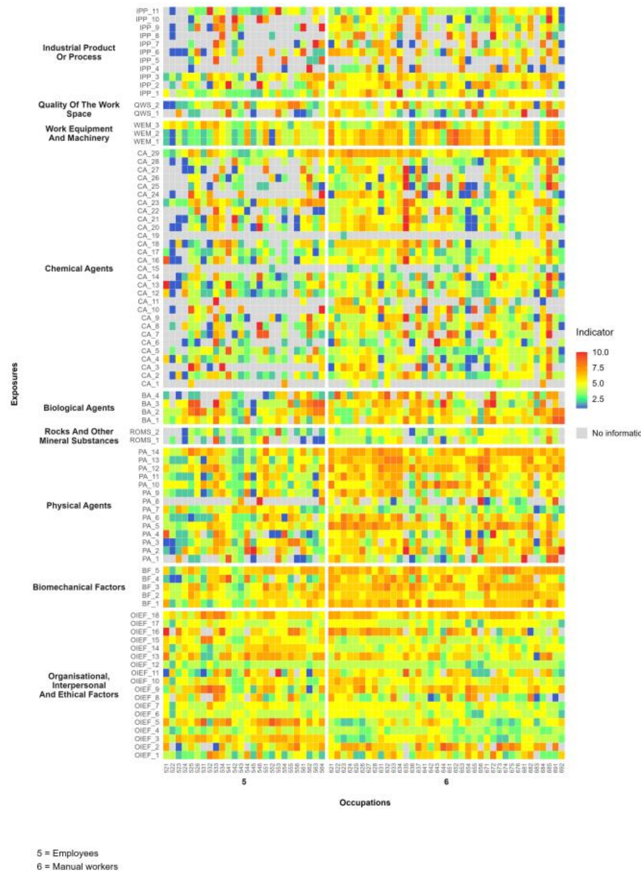


Figure 9. Heatmap of averaged indicators of all occupational groups for occupations 5 to 6 and by exposure.

The heatmap provides a visual representation of occupational exposure profiles, which can be directly linked to the specific tasks performed in each profession. It serves as a valuable tool for identifying high-risk exposure profiles and triggering the necessary protective measures to mitigate occupational hazards.

VI. CASE STUDY TWO: QUALITATIVE ANALYSIS

The integrated dataset facilitates the verification of information concordance across different data sources, thereby enabling the identification of the number of sources reporting on each exposure and disease. This approach makes it possible to visualise both the most and least frequently studied exposures and diseases. When combined with the first case study, this methodology helps to confirm which exposures pose the greatest risks to occupational health—particularly those with high indicator values, studied across multiple sources, and exhibiting a high degree of consistency. To evaluate this concordance, a consistency score has been developed for each exposure and disease based on occupational group.

A. Consistency score construction method for simple exposure or disease

In the data sources, information—excluding missing data—can either confirm or refute the presence of an exposure or disease for a given occupational group. Confirming data sources refer to those that contain values greater than zero for an exposure or disease, indicating that at least one worker in the occupational group has been exposed or affected by the disease. Conversely, refuting data sources contain zero values, signifying no recorded exposure or disease occurrence within the occupational group. The consistency score is determined by comparing the number of sources that confirm an exposure or disease with those that refute it.

If all data sources either confirm or refute a particular exposure or disease, the consistency is categorised as strong, indicating complete agreement among all sources. If an equal number of sources confirm and refute the presence of an exposure or disease, the consistency is categorised as weak, indicating a complete disagreement among all sources. In cases where the number of confirming and refuting sources is unequal, the consistency is categorised as medium. This methodology for calculating the consistency score can be generalised to any number of sources greater than one.

An illustrative case involves skilled craftsmen in building in the “special construction” sector, who are exposed to manual handling of heavy loads and suffer from inner ear diseases. Data on exposure are derived from the Sumer, C2P, and Evrest data sources, while disease-related data come from the AT-MP, MCP, and RNV3P sources (Table III). All exposure-related values in the dataset are greater than zero, and no data sources refute exposure. Consistency is therefore classified as strong. However, in the case of disease, the AT-MP source confirms the presence of disease (positive value), while the MCP and RNV3P sources refute it (zero values). This imbalance results in medium consistency.

An additional example concerns higher-level female secretaries in the “special construction” sector who are exposed to night work. Four data sources provide information on this exposure: MatGene, C2P, Sumer, and Evrest. The first two sources confirm exposure (non-zero values), while the latter two refute it (zero values). As the number of confirming and refuting sources is equal, there is weak consistency in this example.

TABLE III. EXAMPLE OF CONSISTENCY SCORES APPLIED TO DATA

Occupational group	*Exposure / Disease*	*Confirming data sources*	*Refuting data sources*	*Consistency*
Skilled craftsmen in building in the “special construction” activity sector	Manual handling of heavy loads	Sumer: 81 C2P: 37 Evrest: 46	/	Strong
	Diseases of inner ear	AT-MP: 2	MCP: 0 RNV3P: 0	Medium
Higher-level female secretaries in the “special construction” activity sector	Night work	MatGene: 1 C2P: 13	Sumer: 0 Evrest: 0	Weak

Overall, these findings highlight the need for a thorough assessment of the health implications of manually handling heavy loads, as multiple sources indicate that a significant number of workers are exposed to this risk. However, further clarification is required to establish the threshold for night work exposure. Although current methodologies do not enable direct linkage with occupational diseases, safety measures should be reviewed to ensure adequate worker protection.

The consistency scores calculated for each exposure and disease are subsequently visualised using two types of graphical representation: a heatmap and a bar chart. The heatmap depicts the consistency scores of occupational groups for each exposure or disease, while the bar chart illustrates the distribution of consistency scores for exposures across occupational groups based on the number of data sources available.

B. Visualisation of exposure and disease consistency scores in the construction sector

Consistency scores are represented using a blue gradient, where dark blue indicates high consistency and light blue signifies low consistency. The x-axis represents the same occupational groups as in Figure 5 affected by exposures or diseases, while the y-axis displays the exposures or diseases under consideration (see Figure 10).

For example, skilled craftsmen in building in the “special construction” activity sector (group 43_632_1 discussed above) exhibit high consistency scores for manual handling of heavy loads. This occupational group is characterised by a high proportion of strong consistency scores for exposures, with a more moderate representation for diseases. The exposures and occupational groups align with those identified in Case Study 1, where concerns were categorised according to their severity. Notably, the consistency of data sources is strong for most exposures and occupational groups, suggesting that exposures of significant concern are well-documented across all data sources. However, the occupational group 43_486_1 corresponding to foremen and supervisors (excluding administrative supervisors) in maintenance in the “special construction” activity sector lacks data, likely due to the under-recognition of diseases as occupational or caused by workplace exposure. The available information for this group primarily derives from self-reports during medical interviews, revealing a deficiency in how it is monitored.

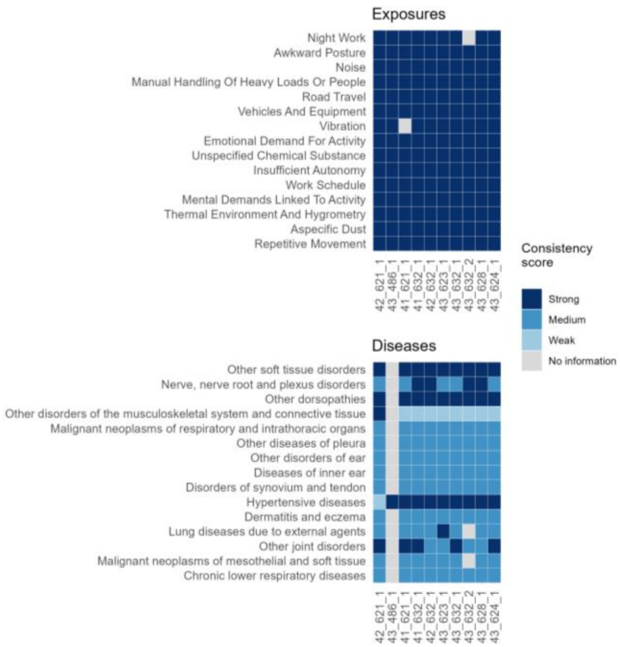


Figure 10. Heatmap of the consistency score by occupational group and by exposure and disease: focus on 10 exposed occupational groups and 15 exposures and diseases.

Overall, consistency scores are weaker for diseases than for exposures. This discrepancy arises from the lower visibility of occupational diseases—some of which, such as hearing loss, are often not officially recognised as occupational illnesses. As a result, such conditions may not be included in the main occupational disease database (AT-MP) but could still be recorded in other databases. This underscores the necessity of integrating diverse disease-related data sources to establish a more comprehensive understanding of occupational health risks.

C. Visualisation of exposure consistency scores in all occupational groups

The bar chart depicted in Figure 11 presents the distribution of consistency score values for exposures across all occupational groups. It is structured into three sections, based on the number of data sources reporting on each exposure. For example, three databases provide information on "wood dust and other plant-based products" (first bar of the graph), while two databases report on "welding fumes" (second bar). The consistency score is only calculated for exposures with information available from at least two data sources. Nearly half of the exposures (n=40) were reported by only one data source and were therefore excluded from the graph. Of the remaining exposures, the majority (n=38) were assessed using data from two sources. Only ten exposures were identified in three sources, and just two exposures were reported across four sources, representing the maximum.

For each section, the x-axis indicates the percentage of occupational groups affected by an exposure, categorised as having high, medium, or low consistency scores. The y-axis represents exposures, grouped according to the eight hierarchical levels of the TEP. The consistency scores are visually represented using four colours: red for high consistency, yellow for medium consistency, green for low consistency, and blue for non-exposed occupational groups. This visualisation highlights the most frequently studied exposures, as represented by a high number of data sources.

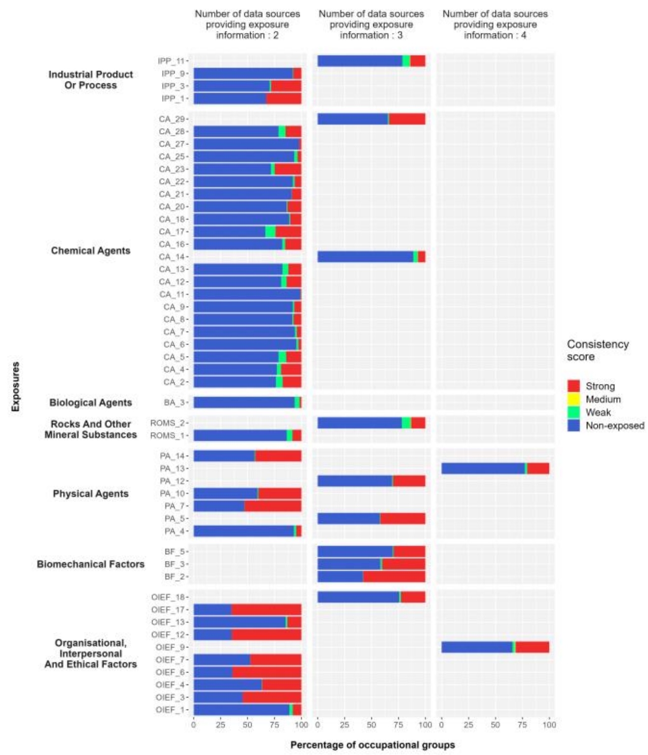


Figure 11. Distribution of exposure consistency scores in all occupational groups.

The distribution of exposure consistency scores reveals that most occupational groups are not exposed to the hazards under study. A substantial proportion of occupational groups with strong consistency scores is observed in relation to organisational, interpersonal, and ethical factors, as well as biomechanical factors and physical agents. A smaller proportion of groups with weak consistency scores is also present, albeit dispersed among different exposure types. Notably, exposures associated with medium consistency scores are nearly invisible on the graph (11 exposures). Importantly, weak consistency does not necessarily reflect poor quality of the data; rather, it may indicate that the information from multiple sources is complementary. This holds true for all exposure types except chemical agents and substances. Additionally, the analysis highlights disparities in data availability for certain exposures, such as biological agents, where only microbiological exposures are documented by more than one data source.

VII. CONCLUSION AND FUTURE WORK

The application of the Occupational Health Data Integration (OBDI) methodology in structuring the dataset has enabled us to implement two distinct analytical approaches for examining occupational health data. To the best of our knowledge, this study represents the first attempt to integrate ten heterogeneous data sources, utilising four domain ontologies. We constructed indicators in order to visualise information as well as underscore occupational groups exposed to multiple hazards and affected by various diseases. Furthermore, consistency scores have been useful in validating these indicators and assessing the complementarity of the data. This methodological approach has facilitated the confirmation of a common body of knowledge, as evidenced by a comparison of our results with those presented in the existing literature.

The two case studies presented here allow for both in-depth analyses at the level of specific occupational groups and broader evaluations encompassing occupations and activity sectors. These case studies offer a comprehensive and precise depiction of exposure and disease profiles across occupational groups, highlighting exposures that impact a significant proportion of the workforce and necessitate ongoing or enhanced safety measures. The visualisation of diseases aids in identifying those profiles that cause the greatest concern, thereby helping in discerning which preventive strategies should be prioritised.

The indicators and consistency scores, although relatively simple and not exhaustive, provide a versatile analytical framework due to the adaptability of their construction methodology. While the case studies serve as illustrative examples of the potential applications of integrated data, we contend that our methodological approach to data integration could be expanded to incorporate additional data sources. An avenue for future research may involve enhancing the first case study by integrating data on the total number of workers within each occupational group, thereby allowing for a more precise estimation of the proportion of exposed and affected workers.

The use of ontologies—particularly the ExO (Exposure Ontology)—provides a powerful framework within which to enhance the integration, interoperability, and value extraction of data in the field of occupational health. National stakeholders in this domain—including public agencies, occupational health services, and research institutes—routinely collect diverse datasets related to occupational exposures, job activities, diseases, and intervention contexts. However, these datasets are often gathered for different purposes, by different actors, and stored in heterogeneous systems, which are rarely connected. This fragmentation severely limits the ability to perform global analyses, conduct longitudinal monitoring, or identify emerging risks.

In this context, the Ontology-Based Data Integration (OBDI) approach, as formalised by Calvanese *et al.*, offers a robust framework for semantically linking these disparate data sources through a shared ontology. The ExO ontology, developed by Mattingly *et al.*, provides an explicit conceptual foundation for describing entities, processes, and relationships in the exposure science domain. It helps standardise terminology, it formalises relationships (*e.g.*, between chemical agents, exposure routes, and occupational activities), and makes data semantically interoperable.

The combined use of ExO and the OBDI approach would enable occupational health stakeholders to answer complex, cross-domain questions that isolated datasets cannot address—for example, identifying multi-source exposures, linking surveillance data with reported occupational diseases, or mapping emerging risks based on various datasets. This approach contributes not only to better data governance, but also—and more importantly—to informed, coordinated decision-making aimed at improving workplace health and safety.

The proposed contextualisation of the data introduces novel perspectives for its application in risk assessment.

For instance, the analysis of correlations between exposure and/or co-exposure indicators could facilitate the identification of key occupational risks, ultimately contributing to the development of a risk assessment tool [26]. The use of these indicators could highlight risks that are little known or studied, and for which preventive measures still need to be developed. Another possibility would be to evaluate or question the effectiveness of measures already put in place for common combined risks to which workers declare themselves to be exposed, in line with the number of accidents at work.

Alternatively, these indicators could enable the creation of worker exposure profiles, consolidating information on all potential exposures and co-exposures associated with a given occupation [21]. Such profiles could be subsequently employed in order to anticipate future occupational diseases and implement occupation-specific safety measures. When calculating the indicator, the tasks performed, as well as biological characteristics such as sex should also be considered in order to better represent the diversity of work situations and workers, and to compare differences in exposure with a view to improving individual safety.

Although our article is focused on the field of occupational health, the proposed method of use and case studies could be generalised to the broader healthcare sector. Current research in the health sector has increasingly focused on the contextualisation of heterogeneous data to establish a common semantic framework aimed at enhancing knowledge sharing [27], [28]. Generalising our approach could facilitate the development of targeted responses to both existing and emerging health concerns through specialised tools and analytical queries. Moreover, it would promote interdisciplinary collaboration and facilitate the generation of knowledge grounded in practical, real-world scenarios.

ACKNOWLEDGEMENTS

The authors would like to thank the various healthcare organisations that agreed to participate in the DataPOST project and to share their data. The work presented was based on data from the Dares Sumer survey, data supplied by the GIS EVREST and made available on 01/08/2024, and data provided by the Anses, CNAM, SPF, and INRS.

The authors would also like to thank E. Algava and M. Duval (Dares), L. Meunier (CNAM), C. Pilorget, J. Chatelot, and J. Homère (SPF), C. Nisse and A. Aachimi (Anses), and L. Rollin and A. Leroyer (Evrest) for their expertise.

REFERENCES

[1] C. Barbey, N. Bonvallot, M. Smaïl-Tabbone, and F. Clerc, 'Ontology-Based Integration of Occupational Health Data: Method and Case Studies', *DATA Anal.*, 2024.

[2] C. Barbey, N. Bonvallot, and F. Clerc, 'Health Outcomes Related to Multiple Exposures in Occupational Settings: A Review', *Saf. Health Work*, vol. 15, no. 4, pp. 382–395, Dec. 2024, doi: 10.1016/j.shaw.2024.10.004.

[3] L. Rollin *et al.*, 'What complementarity for french occupational health databases (Evrest, MCP, Sumer, RNV3P): an example from the women home help workers?', *Arch. Mal. Prof. Environ.*, vol. 82, no. 3, pp. 261–276, May 2021, doi: 10.1016/j.admp.2020.11.002.

[4] L. Rollin *et al.*, 'Les conducteurs de poids lourd, transport en commun et livreurs sont-ils à risque d'affections du membre supérieur ?', *Arch. Mal. Prof. Environ.*, vol. 85, no. 6, p. 102765, Nov. 2024, doi: 10.1016/j.admp.2024.102765.

[5] R. Dandan, S. Despres, and J. Nobecourt, 'OAFE: An Ontology for the Description of Elderly Activities', in *2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, Nov. 2018, pp. 396–403. doi: 10.1109/SITIS.2018.00068.

[6] D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati, 'Ontology-Based Data Access and Integration', in *Encyclopedia of Database Systems*, 2017, pp. 1–7. doi: 10.1007/978-1-4899-7993-3_80667-1.

[7] A. Z. E. Qutaany, O. M., and A. H., 'A Mapping Approach for Fully Virtual Data Integration System Processes', *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 12, 2018, doi: 10.14569/IJACSA.2018.091216.

[8] H. Zhang *et al.*, 'An ontology-guided semantic data integration framework to support integrative data analysis of cancer survival', *BMC Med. Inform. Decis. Mak.*, vol. 18, no. S2, p. 41, Jul. 2018, doi: 10.1186/s12911-018-0636-4.

[9] R. Dandan, 'Intégration de données guidée par une ontologie : Application à l'assistance personnalisée aux personnes âgées', Université Paris-Nord - Paris XIII, 2022. [Online]. Available: https://theses.hal.science/tel-03945528

[10] C. J. Mattingly, T. E. McKone, M. A. Callahan, J. A. Blake, and E. A. C. Hubal, 'Providing the Missing Link: the Exposure Science Ontology ExO', *Environ. Sci. Technol.*, vol. 46, no. 6, pp. 3046–3053, Mar. 2012, doi: 10.1021/es2033857.

[11] J. Bloch *et al.*, 'National Network for Monitoring Prevention of a Occupational Disease (RNV3P) - 2022 Annual Report', Oct. 2023.

[12] J. A. Hirsch *et al.*, 'ICD-10: History and Context', *Am. J. Neuroradiol.*, vol. 37, no. 4, pp. 596–599, Apr. 2016, doi: 10.3174/ajnr.A4696.

[13] C. Brossard and I. Falinower, 'Salariés déclarés exposés à des risques de pénibilité en 2016 : portrait':, *Retraite Société*, vol. N° 77, no. 2, pp. 139–163, Oct. 2018, doi: 10.3917/rs1.077.0139.

[14] G. Mater, C. Paris, and J. Lavoué, 'Descriptive analysis and comparison of two French occupational exposure databases: COLCHIC and SCOLA', *Am. J. Ind. Med.*, vol. 59, no. 5, pp. 379–391, May 2016, doi: 10.1002/ajim.22569.

[15] F. Bonthoux and N. Schmitt, 'Colphy : base de données dédiée aux nuisances physiques', *Hyg. Sécurité Trav.*, no. 270, pp. 61–65, 2023.

[16] J. Févotte *et al.*, 'Matgéné: A Program to Develop Job-Exposure Matrices in the General Population in France', *Ann. Occup. Hyg.*, vol. 55, no. 8, pp. 865–878, Sep. 2011, doi: 10.1093/annhyg/mer067.

[17] Assurance Maladie - Risques professionnels, 'Rapport annuel 2023 de l'Assurance Maladie - Risques professionnels', Paris, Dec. 2024.

[18] Z. Zhu, 'Exoskeletons for manual material handling – A review and implication for construction applications', *Autom. Constr.*, 2021.

[19] J. J. Devereux, I. G. Vlachonikolis, and P. W. Buckle, 'Epidemiological study to investigate potential interaction between physical and psychosocial factors at work that may increase the risk of symptoms of musculoskeletal disorder of the neck and upper limb', *Occup. Environ. Med.*, vol. 59, no. 4, pp. 269–277, Apr. 2002, doi: 10.1136/oem.59.4.269.

[20] L. L. Andersen, J. Vinstrup, E. Sundstrup, S. V. Skovlund, E. Villadsen, and S. V. Thorsen, 'Combined ergonomic exposures and development of musculoskeletal pain in the general working population: A prospective cohort study', *Scand. J. Work. Environ. Health*, vol. 47, no. 4, pp. 287–295, May 2021, doi: 10.5271/sjweh.3954.

[21] C. Fourneau *et al.*, 'The French 2016-2020 National Occupational Health Plan: a better understanding of multiple exposures', *Environ. Risques Santé*, vol. 20, no. 4, pp. 377–382, Aug. 2021, doi: 10.1684/ers.2021.1570.

[22] T.-H.-Y. Nguyen, M. Bertin, J. Bodin, N. Fouquet, N. Bonvallot, and Y. Roquelaure, 'Multiple Exposures and Coexposures to Occupational Hazards Among Agricultural Workers: A Systematic Review of Observational Studies', *Saf. Health Work*, vol. 9, no. 3, pp. 239–248, Sep. 2018, doi: 10.1016/j.shaw.2018.04.002.

[23] M. Ekmekci and S. Yaman, 'Occupational health and safety among farmers: a comprehensive study in Central Anatolia, Turkey', *BMC Public Health*, vol. 24, no. 1, p. 2732, Oct. 2024, doi: 10.1186/s12889-024-20249-7.

[24] R. Anders, A. Frapsauce, C. Sauvezon, and D. Gilibert, 'Police officer occupational health: a model of organizational constraints, trauma exposure, perceived resources, and

agency', *J. Occup. Med. Toxicol.*, vol. 19, no. 1, p. 46, Dec. 2024, doi: 10.1186/s12995-024-00444-3.

[25] S. St. Louis, N. A. Frost, C. E. Monteiro, and J. Trapassi Migliaccio, 'Occupational Hazards in Corrections: The Impact of Violence and Suicide Exposures on Officers' Emotional and Psychological Health', *Crim. Justice Behav.*, vol. 50, no. 9, pp. 1361–1379, Sep. 2023, doi: 10.1177/00938548231177710.

[26] A. J. Williams, J. C. Lambert, K. Thayer, and J.-L. C. M. Dorne, 'Sourcing data on chemical properties and hazard data from the US-EPA CompTox Chemicals Dashboard: A practical guide for human risk assessment', *Environ. Int.*, vol. 154, p. 106566, Sep. 2021, doi: 10.1016/j.envint.2021.106566.

[27] R. R. Boyles, A. E. Thessen, A. Waldrop, and M. A. Haendel, 'Ontology-based data integration for advancing toxicological knowledge', *Curr. Opin. Toxicol.*, vol. 16, pp. 67–74, Aug. 2019, doi: 10.1016/j.cotox.2019.05.005.

[28] R. R. Rao, K. Makkithaya, and N. Gupta, 'Ontology based semantic representation for Public Health data integration', in *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, Mysore, India: IEEE, Nov. 2014, pp. 357–362. doi: 10.1109/IC3I.2014.7019701.

ANNEXES

Annex A. Correspondence between codes and labels for activity sector divisions.

A - Agriculture, forestry, and fishing	
01	Crop and animal production, hunting, and related service activities
02	Forestry and logging
03	Fishing and aquaculture
B - Mining and quarrying	
05	Mining of coal and lignite
06	Extraction of crude petroleum and natural gas
07	Mining of metal ores
08	Other mining and quarrying
09	Mining support service activities
C - Manufacturing	
10	Manufacture of food products
11	Manufacture of beverages
12	Manufacture of tobacco products
13	Manufacture of textiles
14	Manufacture of wearing apparel
15	Manufacture of leather and related products
16	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
17	Manufacture of paper and paper products
18	Printing and reproduction of recorded media
19	Manufacture of coke and refined petroleum products
20	Manufacture of chemicals and chemical products
21	Manufacture of basic pharmaceutical products and pharmaceutical preparations
22	Manufacture of rubber and plastic products
23	Manufacture of other non-metallic mineral products
24	Manufacture of basic metals
25	Manufacture of fabricated metal products, except machinery and equipment
26	Manufacture of computer, electronic, and optical products
27	Manufacture of electrical equipment
28	Manufacture of machinery and equipment
29	Manufacture of motor vehicles, trailers, and semi-trailers
30	Manufacture of other transport equipment
31	Manufacture of furniture
32	Other manufacturing
33	Repair and installation of machinery and equipment
D - Electricity, gas, steam, and air conditioning supply	
35	Electricity, gas, steam, and air conditioning supply
E - Water supply; sewerage, waste management, and remediation activities	
36	Water collection, treatment, and supply
37	Sewerage
38	Waste collection, treatment, and disposal activities; materials recovery
39	Remediation activities and other waste management services
F - Construction	
41	Construction of buildings
42	Civil engineering
43	Specialised construction activities
G - Wholesale and retail trade; repair of motor vehicles and motorcycles	
45	Wholesale and retail trade and repair of motor vehicles and motorcycles
46	Wholesale trade, except of motor vehicles and motorcycles
47	Retail trade, except of motor vehicles and motorcycles
H - Transportation and storage	
49	Land transport and transport via pipelines
50	Water transport
51	Air transport
52	Warehousing and support activities for transportation
53	Postal and courier activities
I - Accommodation and food service activities	
55	Accommodation
56	Food and beverage service activities
J - Information and communication	
58	Publishing activities
59	Motion picture, video, and television programme production, sound recording, and music publishing activities
60	Programming and broadcasting activities
61	Telecommunications
62	Computer programming, consultancy, and related activities
63	Information service activities
K - Financial and insurance activities	
64	Financial service activities, except insurance and pension funding
65	Insurance, reinsurance, and pension funding, except compulsory social security
66	Activities auxiliary to financial services and insurance activities
L - Real estate activities	
68	Real estate activities
M - Professional, scientific, and technical activities	
69	Legal and accounting activities
70	Activities of head offices; management consultancy activities
71	Architectural and engineering activities; technical testing and analysis
72	Scientific research and development
73	Advertising and market research
74	Other professional, scientific, and technical activities
75	Veterinary activities
N - Administrative and support service activities	
77	Rental and leasing activities
78	Employment activities
79	Travel agency, tour operator, and other reservation service and related activities
80	Security and investigation activities
81	Services to buildings and landscape activities
82	Office administrative, office support, and other business support activities
O - Public administration and defence; compulsory social security	
84	Public administration and defence; compulsory social security
P - Education	
85	Education
Q - Human health and social work activities	
86	Human health activities
87	Residential care activities
88	Social work activities without accommodation
R - Arts, entertainment, and recreation	
90	Creative, arts, and entertainment activities
91	Libraries, archives, museums, and other cultural activities
92	Gambling and betting activities
93	Sports activities and amusement and recreation activities
S - Other service activities	
94	Activities of membership organisations
95	Repair of computers and personal and household goods
96	Other personal service activities
T - Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use	
97	Activities of households as employers of domestic personnel
98	Undifferentiated goods- and services-producing activities of private households for own use
U - Activities of extraterritorial organisations and bodies	
99	Activities of extraterritorial organisations and bodies

Annex B. Exposure code details.

IPP - Industrial product or process	
IPP_11	Wood dust and other plant-based product
IPP_10	Steel and metal
IPP_9	Product released in welding processes
IPP_8	Product released in foundry processes
IPP_7	Product of organic origin
IPP_6	Product of inorganic origin
IPP_5	Pharmaceutical products
IPP_4	Nanomaterial nanoparticle
IPP_3	Miscellaneous industrial equipment and products
IPP_2	Construction products
IPP_1	Non-specific dust
QWS - Quality of the work space	
QWS_2	Constrained workspace
QWS_1	Clean room and workspace with a special situation
WEM - Work equipment and machinery	
WEM_3	Vehicles and equipment
WEM_2	Machine tool
WEM_1	Construction machine
CA - Chemical agents	
CA_29	Unspecified chemical substance
CA_28	Transition metal (cadmium, chromium, nickel, cobalt, etc.)
CA_27	Sulphoxide and sulphone
CA_26	Sulphonic acid and thioacid
CA_25	Phenol and derivatives
CA_24	Organic metal compounds
CA_23	Non-metal
CA_22	Nitrile cyanate, isocyanate, and cyanurate
CA_21	Natrium, potassium, lithium, and other alkaline metals
CA_20	Metalloid
CA_19	Lanthanide and rare earths
CA_18	Lactone and lactam
CA_17	Hydrocarbons and drift
CA_16	Halogen
CA_15	Glycol
CA_14	Formaldehyde and other aldehydes
CA_13	Ether thioether and derivatives
CA_12	Ester
CA_11	Epoxy
CA_10	Carboxylic acid salt
CA_9	Carboxylic acid and peracid
CA_8	Beryllium, magnesium, and other alkaline earth metals
CA_7	Amine imine and drift
CA_6	Amide sulphonamide phosphoramide imide and thiuram
CA_5	Aluminium, lead and other poor metal
CA_4	Alcohol and polyalcohol and derivatives
CA_3	Actinide
CA_2	Acetone and other cetone, quinone, cetene, and derivatives
CA_1	Acetal and derivatives
BA - Biological agents	
BA_4	Vegetal
BA_3	Microbiological
BA_2	Biological agents
BA_1	Animal
ROMS - Rocks and other mineral substances	
ROMS_2	Silica and other sedimentary rocks
ROMS_1	Asbestos and other silicate minerals
PA - Physical agents	
PA_14	Walk on the same level
PA_13	Vibration
PA_12	Thermal environment and hygrometry
PA_11	Static field
PA_10	Road travel
PA_9	Radiation and electromagnetic fields
PA_8	Pressure above atmospheric pressure
PA_7	Other lighting and visual constraints
PA_6	Non-ionising radiation
PA_5	Noise
PA_4	Ionising radiation
PA_3	Insufficient ventilation
PA_2	Constraint linked to humidity
PA_1	Artificial light
BF - Biomechanical factors	
BF_5	Repetitive movement
BF_4	Other biomechanical factor
BF_3	Manual handling of heavy loads or people
BF_2	Awkward posture
BF_1	Another position
OIEF - Organisational, interpersonal, and ethical factors	
OIEF_18	Workstation
OIEF_17	Work schedule
OIEF_16	Work experience contrary to its principles
OIEF_15	Teleworking
OIEF_14	Special working arrangements
OIEF_13	Quality of working relationships
OIEF_12	Quality of life prevented
OIEF_11	Other organisational, relational, and ethical factors
OIEF_10	Other functional organisation of activity
OIEF_9	Night work
OIEF_8	Methods that increase psychosocial risk
OIEF_7	Mental demands linked to activity
OIEF_6	Insufficient autonomy
OIEF_5	External violence
OIEF_4	Experience of overwork or underwork
OIEF_3	Emotional demand for activity
OIEF_2	Company-related bonus factor
OIEF_1	Business travel

Annex C. Correspondence between codes and labels for intermediate professions.

1 - Farm operators	
111	Smallholding farmers
122	Activities similar to farmers on medium-sized holdings
131	Farmers on large holdings
2 - Craftsmen, shopkeepers, and company directors	
210	Craftsmen in general
211	Craftsmen in building, public works, parks, and gardens
212	Craftsmen in metalworking, mechanics, electromechanics, and electrical equipment
213	Craftsmen in textiles, clothing, and leather
214	Craftsmen in furniture, woodworking, and other manufacturing
215	Food craftsmen
216	Craftsmen in repairs and maintenance
217	Craftsmen in other services
218	Craftsmen-like
219	Craftsmen’s family helpers
220	Retailers and similar in general
221	Small and medium-sized wholesalers (0 to 9 employees)
222	Small and medium-sized food retailers (0 to 9 employees)
223	Small and medium-sized specialist retailers (by field) (0 to 9 employees)
224	Operators of cafés, restaurants, and hotels (0 to 9 employees)
225	Retail intermediaries (0 to 9 employees)
226	Agents of insurance, transport, and tourism (0 to 9 employees)
227	Other service providers (0 to 9 employees)
231	Managers of large companies (500 employees or more)
232	Managers of medium-sized companies (50 to 499 employees)
233	Managers of companies with 10 to 49 employees
3 - Executives and higher intellectual professions	
311	Self-employed health professionals
312	Self-employed legal and technical professions
313	Self-employed caregivers
331	Public service: Executive staff
332	Public service: Engineers and senior technical staff
333	Public service: Administrative staff (excluding teaching and heritage)
334	Public service: Army and Gendarmerie officers
335	Public service: Political and trade union representatives
341	Professors and scientific occupations in secondary education
342	Professors and scientific occupations in higher education and public research
343	Professors and scientific occupations in educational and vocational guidance
344	Professors and scientific occupations in health
351	Information, arts, and entertainment professions: Civil service in documentation, heritage
352	Information, arts, and entertainment professions: Journalism, creative writing
353	Information, arts, and entertainment professions: Managers in the press, publishing, audiovisual, and entertainment industries
354	Information, arts, and entertainment professions: Artists
371	Executive managers (large companies with 500 or more employees)
372	Administrative and financial specialists
373	Other administrative and financial managers
374	Commercial administration, commercial function
375	Administrative and commercial managers in advertising, public relations, communication
376	Administrative and commercial managers in banking, insurance and real estate, financial market managers
377	Administrative and commercial managers in hotels, restaurants
380	Engineers and technical managers: Executive staff (large companies)
381	Engineers and technical managers in agriculture, water, and forestry
382	Engineers and technical managers in building, public works
383	Engineers and technical managers in electricity, electronics
384	Engineers and technical managers in mechanics, metalworking
385	Engineers and technical managers in processing industries (agri-food, chemicals, metallurgy, and heavy materials)
386	Engineers and technical managers in other industries (printing, soft materials, furniture and wood, energy)
387	Engineers and technical managers in related production functions: Industrial purchasing, logistics, methods, quality control, maintenance (excluding IT), environment, etc.
388	Engineers and technical managers in IT, telecommunications
389	Engineers and technical managers in transport (excluding logistics)
4 - Intermediate occupations	
421	School teachers and similar staff in nurseries and elementary schools
422	Other school teachers and similar staff
423	School teachers and similar staff in continuing education
424	School teachers and similar staff in sports education and professional sport
425	School teachers and similar staff in documentation and heritage
431	Intermediate health and social work professions: Nurses, midwives, and related professions
432	Intermediate health and social work professions: Physiotherapists and rehabilitation specialists
433	Intermediate health and social work professions: Medical technicians and medical equipment specialists
434	Intermediate health and social work professions: Specialists in socio-educational intervention
435	Intermediate health and social work professions: Specialists in socio-cultural and leisure activities
441	Clergy and religious
451	Intermediate administrative professions in the civil service: Administrative staff
452	Intermediate administrative professions in the civil service: Police and military personnel
461	Intermediate administrative and commercial professions: Higher-level secretaries, corporate administrative services supervisors
462	Intermediate administrative and commercial professions: Sales shop managers, purchasing function, sales administration
463	Intermediate administrative and commercial professions: Sales force technicians, sales representatives
464	Intermediate administrative and commercial professions: Advertising, public relations, communication
465	Intermediate administrative and commercial professions: Artistic expression, entertainment
466	Intermediate administrative and commercial professions: Transport, tourism
467	Intermediate administrative and commercial professions: Technical and commercial services in banking, insurance, and social security organisations
468	Intermediate administrative and commercial professions: Hotels and restaurants
471	Technical staff in agriculture, water, and forestry
472	Technical staff in building, public works
473	Technical staff in electricity, electromechanics, and electronics
474	Technical staff in mechanics, metalworking
475	Technical staff in processing industries (agri-food, chemicals, metallurgy and heavy materials)
476	Technical staff in other industries (printing, soft materials, furniture and wood)
477	Technical staff in related production functions: logistics, maintenance (excluding IT), environment
478	Technical staff in informatics and telecommunications
479	Technical staff: Experts and public research technicians
480	Foremen and supervisors (excluding administrative supervisors) in agriculture, water and forestry, maritime professions
481	Foremen and supervisors (excluding administrative supervisors) in building, public works
482	Foremen and supervisors (excluding administrative supervisors) in electricity, electronics
483	Foremen and supervisors (excluding administrative supervisors) in mechanical engineering, metalworking
484	Foremen and supervisors (excluding administrative supervisors) in processing industries (agri-food, chemicals, metallurgy, heavy materials)
485	Foremen and supervisors (excluding administrative supervisors) in other industries (printing, flexible materials, furniture and wood, energy, water)
486	Foremen and supervisors (excluding administrative supervisors) in maintenance, new works
487	Foremen and supervisors (excluding administrative supervisors) in warehousing, storage, and handling
488	Foremen and supervisors (excluding administrative supervisors) in restaurants
5 - Employees	
521	Civil servants and service agents in the public service: Post Office and France Télécom employees (public status)
522	Civil servants and service agents in the public service: Specialised tax, treasury, and customs agents
523	Civil servants and service agents in the public service: Administrative assistants in the public service (including teaching)
524	Civil servants and service agents in the public service: Administrative officers in the public service (including teaching)
525	Civil servants and service agents in the public service: Service agents in the public service (including teaching) and similar (hospitals and private clinics)
526	Civil servants and service agents in the public service: Caregivers and related professions in the public or private sector
531	Police and military: Police and prison guards

532	Police and military: Army
533	Police and military: Firefighters, nature, and heritage conservation officers
534	Police and military: Security and surveillance officers
541	Company administrative employees in reception and information
542	Company administrative employees in secretarial and typing duties
543	Company administrative employees in accounting and administrative services
544	Company administrative employees in information technology
545	Company administrative employees in banking, insurance, and social security technical and commercial services
546	Company administrative employees in transport, tourism
551	Commercial employees in procurement, labelling
552	Commercial employees: Cashiers
553	Commercial employees in non-specialised sales
554	Commercial employees in specialised sales (by field)
555	Commercial employees in mail order, telesales
556	Commercial employees in wholesale of capital and intermediate goods
561	Direct services to individuals in hotels, cafés, restaurants
562	Direct services to individuals in personal care services
563	Direct services to individuals in social intervention and domestic help
564	Direct services to individuals (Miscellaneous)
6 - Manual workers	
621	Skilled industrial workers in building, public works, quarries, extraction
622	Skilled industrial workers in electricity, electronics
623	Skilled industrial workers in metal working
624	Skilled industrial workers in mechanics, mechanical engineering
625	Skilled industrial workers in food, chemical, and related industries (chemicals, plastics, pharmaceuticals, water, energy)
626	Skilled industrial workers in processing industries (metallurgy, glass production, building materials)
627	Skilled industrial workers in other industries (textiles, haberdashery, clothing, industrial leatherworking, woodworking, furniture, paper and cardboard, printing)
628	Skilled industrial workers in maintenance, industrial equipment maintenance, adjustment, laboratory work
631	Skilled craftsmen in gardening
632	Skilled craftsmen in building
633	Skilled craftsmen in electricity, electronics
634	Skilled craftsmen in mechanics, metalworking
635	Skilled craftsmen in textiles, clothing, leather
636	Skilled craftsmen in food, catering
637	Skilled craftsmen (miscellaneous)
641	Lorry drivers
642	Taxi and private car drivers
643	Delivery drivers, couriers
644	Drivers related to environmental purposes (waste collection, sanitation, cleaning)
651	Skilled workers in handling, warehousing, and transport: Conductors of heavy lifting and shunting equipment
652	Skilled workers in handling, warehousing, and transport: Skilled handling workers, forklift truck drivers, forklift truck operators
653	Skilled workers in handling, warehousing, and transport: Warehousemen
654	Skilled workers in handling, warehousing, and transport: Qualified drivers of guided transport vehicles
655	Skilled workers in handling, warehousing, and transport: Other qualified transport workers
656	Skilled workers in handling, warehousing, and transport: Merchant seamen, captains, and helmsmen on inland waterways
671	Unskilled industrial workers in building, public works, quarries, extraction
672	Unskilled industrial workers in electricity, electronics
673	Unskilled industrial workers in forging, metalworking, mechanics
674	Unskilled industrial workers in processing industries (chemicals, pharmaceuticals, plastics, food processing, metal processing, glass, building materials)
675	Unskilled industrial workers in other industries (textiles, clothing, leather, wood, furniture, paper and cardboard, printing, press, publishing)
676	Unskilled industrial workers in handling, sorting, packaging, shipping, miscellaneous
681	Unskilled craftsmen in building
682	Unskilled craftsmen in mechanics
683	Unskilled craftsmen in power supply
684	Unskilled craftsmen in cleaning, sanitation, waste treatment
685	Other unskilled craftsmen
691	Agricultural workers in agriculture, forestry
692	Agricultural workers in fishing, aquaculture

Detection of Changes in Lateral Weight Shift During Gait Improvement of Patients via Image Analysis of Frontal-Direction Imaging Using MediaPipe

Yasutaka Uchida
Dept. of Life Science,
Teikyo University of Science
Tokyo, Japan
e-mail: uchida@ntu.ac.jp

Tomoko Funayama
Dept. of Occupational Therapy
Teikyo University of Science
Yamanashi, Japan
e-mail: funayama@ntu.ac.jp

Eiichi Ohkubo
Dept. of Life Science,
Teikyo University of Science
Tokyo, Japan
e-mail: ohkubo@ntu.ac.jp

Daisuke Souma
Dept. of Rehabilitation
Isogo Central Hospital
Yokohama, Japan
e-mail: rifles060920@gmail.com

Yasunori Fujimori
Dept. of Rehabilitation
Seirei Yokohama Hospital
Yokohama, Japan
e-mail: yasu6622@outlook.jp

Yoshiaki Kogure
Professor Emeritus
Teikyo University of Science
Tokyo, Japan
e-mail: kogure@ntu.ac.jp

***Abstract*—Obtaining numerical skeletal data is important for objectively assessing a patient's gait status. Using MediaPipe on images captured from the frontal plane—a method not limited by the imaging-facility size—we use skeletal data to evaluate walking changes caused by the rehabilitation of three patients over a period of approximately one month. To reduce errors due to differences in participant picture size, two methods are examined: one based on waist width and the other on shoulder width. Using standardized data, we examine the changes in the nose position, left and right ankle heights, and left and right shoulder slopes, as well as the differences between toes and heels, and between the left and right elbow widths. Additionally, the appearance of each part of the body during walking is examined via spectrograms. Despite the few participants, the findings suggest that not only the stride length and foot speed but also the blurring of the nose position, left and right elbow positions, and ankle and knee can be used as indicators to assess whether weight is being shifted smoothly during walking as well as to evaluate the degree of recovery from rehabilitation.**

Keywords—MediaPipe; rehabilitation evaluation; digital health; frontal-direction image; gait analysis; lateral weight shift.

I. INTRODUCTION

Assessing changes in a patient's condition is crucial for rehabilitation [1]-[4]. Gait assessments in clinical settings reveal considerable potential health status and predictive information. Quantitative instrumented gait analysis is recommended for clinical gait assessment; however, its use is currently limited. Owing to recent advances in machine-learning research, studies pertaining to rehabilitation recovery are increasing rapidly. According to previous studies, various sensors are used to measure the time required to perform a defined exercise, and the data are used in machine-learning methods such as k-nearest neighbor approximation, support vector machine, random forest, and logistic regression [5].

Spatiotemporal parameters during gait are considered effective for quantifying gait performance and determining the state of physical functions. Measurement units are not limited by the measurement space as they do not require a pre-installed three-dimensional motion-capture system, for instance the system released by Vicon Motion Systems Ltd. UK, which is used as the de facto standard; however, this standard must be validated [6][7].

Kinect for Windows v1, which is released by Microsoft, features a red-green-blue camera for color videos and an infrared (IR) emitter. The camera allows depth measurement when the baseline between the camera and projector is known. Meanwhile, Kinect for Windows v2 offers improved skeleton tracking. Azure Kinect DK, which was released in 2019, integrates artificial-intelligence applications. The potential for realizing clinical applications using the Azure Kinect camera, which is continuously being improved, is currently being investigated [8]-[12].

Recent advances in machine learning and other technologies have enabled skeletal recognition in software such as OpenPose [13]-[17] without using IR cameras such as the Kinect. OpenPose is currently used for knee- and ankle-motion analyses because it can estimate whole-body skeletons and human postures. MediaPipe [18]-[29] supports various frameworks and uses video cameras and images captured by smartphone cameras for analysis. It offers advantages such as the use of a high-performance graphics processing units through Google Colaboratory. However, no-code programming is currently possible, although it requires understanding regarding the source code. Therefore, applications such as reporting gait analysis are rare. Additionally, because the rehabilitation area is required for video recording, analysis from the sagittal-plane direction is challenging, and only a few implementation cases have been reported.

At a conference held last year, we reported the possibility of using MediaPipe to analyze images captured from the frontal plane based on the stride length and ankle angle. Additionally, we measured the gait of participants who underwent physical and occupational therapy training and verified the effectiveness of walking aids in participants who required hospitalization. The effectiveness of the walking aids was verified as follows: the dispersion of the nose position in the left and right directions was used as an index of body shaking during walking; the tilt of the shoulders, hips, and neck during walking was calculated using MediaPipe data to determine balance; and the time variation of these data was used as the basis for discrete Fourier spectrum decomposition and heel spectrograms. The effect of the walking aids on the participants' gait was demonstrated from multiple perspectives.

In this study, we use a method reported in the literature for participants under different conditions; subsequently, we examine the differences and attempt to improve the accuracy. The remainder of this paper is organized as follows: Section II presents the experimental conditions. In Section III, we present the experimental results obtained via characteristic observations performed by participants undergoing rehabilitation, as previously reported, toward other participants and analyze them from a new perspective. Section IV discusses the obtained results. The proposed method reduces the effect of the distance from the camera through hip or shoulder-width standardization. Changes in gait are confirmed by determining the nose position, height at the feet, tilt of the right and left shoulders, toe and heel widths, and elbow width during gait, as well as based on spectrograms. Section V concludes the paper.

This study was approved by the Ethics Committee on Research with Humans as Participants of the Teikyo University of Science. The participants received written informed consent from the physical therapist at the hospital where they were admitted.

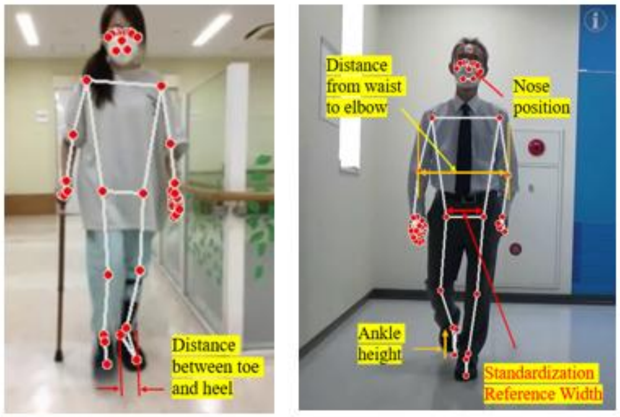
II. EXPERIMENTS

Video recordings of walking conditions on the ORPHE ANALYTICS screen were captured using a snipping tool and analyzed using MediaPipe. During the measurements, walking at a distance of 3 m from the front was recorded using a smartphone camera. Participant A was a female in her 40s. She was diagnosed with right capsular hemorrhage, and her disabilities included left hemiplegia and severe sensory impairment. Prior to the onset of stroke, she was able to perform daily activities independently, was employed full-time five days a week, and commuted to work using public transportation. She was transferred to the hospital for convalescent rehabilitation approximately two weeks after the onset of stroke. Physical therapy (120 min) and occupational therapy (60 min) were provided to the patient, and gait training was conducted during physical therapy. The first image was captured 72 d after stroke onset, and the second image was captured on the 109th day. The patient was discharged from the hospital 4 d after the video recording, and rehabilitation was performed on an outpatient basis (Participant A). When she was discharged from the hospital, she walked outdoors with the assistance of a T-cane and short leg brace.

Participant B was a right-handed man in his 70s. He was diagnosed with left atherothrombotic cerebral infarction, right hemiplegia, and mild deep sensory dullness. His medical history included a cervical spinal-cord injury (difficulty in lifting the right upper limb), lung-cancer surgery, and pharyngeal cancer (currently undergoing radiation therapy). Prior to the onset of cerebral infarction, his daily activities included crawling indoors. He was transferred to Isog Central Hospital 18 d after the onset of stroke for rehabilitation and was underwent 60–80 minutes of physical therapy and 60 min of occupational therapy seven times a week, during which he underwent walking training.

Participant C was a right-handed man in his 60s. He was diagnosed with left thalamic hemorrhage and right hemiplegia, and anesthesia. His medical history included chronic heart failure, hypertension, chronic kidney disease, schizophrenia, colonic polyps, and gastroesophageal reflux disease. Before the onset of cerebral hemorrhage, the patient lived alone and was able to perform daily activities independently, including outdoor activities. He was transferred to Isogo Central Hospital on the day of the onset for blood-pressure management and follow-up. Subsequently, 29 days after the onset, he was transferred to our hospital's rehabilitation ward for continued rehabilitation, where he received 120 min of physical therapy and 60 min of occupational therapy seven times a week, including gait training.

For reference, a participant without walking disabilities (Participant D in this study) videotaped a male participant in his 60s at Teikyo University of Science. The video of Participant D's analysis was captured using a camera.



(a) Participant A (b) Participant D

Figure 1. Definition of evaluation parameters.

As shown in Figure 1, the waist width was used as the standard, and the coordinates of each part were assigned and normalized using this value. This was performed to address measurement errors in parameters during walking, such as the stride length, caused by the small image of the participant at the start of walking. Additionally, it enabled one to determine the stride length and foot speed more easily based on the measurement results. The values shown in the results were defined as the distance between the toes and heel, the nose position, the distance from the waist to the elbow, and the ankle height, as shown in Figures 1(a) and 1(b).

We used a wireless smart insole (FEELSOLE®) equipped with pressure sensors, which enabled four parts (the toe, heel, inside, and outside) of each foot to be measured, thus resulting in eight parts for both feet. The insoles were calibrated before use. Calibration was performed four times: with no pressure or feet in shoes, standing on both feet, and standing on one foot on each side. The sampling frequency was set to 50 Hz. Data were stored on the cloud using ORPHE ANALYTICS and downloaded in the CSV format [30].

III. Experimental Results

A. *Height of left and right ankles*

When the images were captured from the front, owing to the features of MediaPipe, the z-axis values increased as the participant approached the camera. The z-axis values of the ankles and other parts of the body at the start of walking were normalized by the width of the hips to avoid ambiguity in the gait conditions, such as the stride length. Figure 2 shows the left and right ankle heights normalized by the hip width for the participant 72 d after onset. Figure 3 shows the ankle height of the same participant 109 d after the onset, normalized by the hip width. In Figures 2 and 3, the horizontal axis shows the time in seconds, and the vertical axis shows the ankle position from the normalized waist.

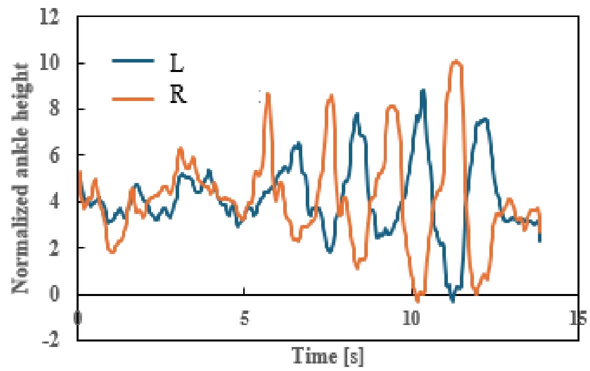


Figure 2. Normalized ankle height during gait 72 d after onset.

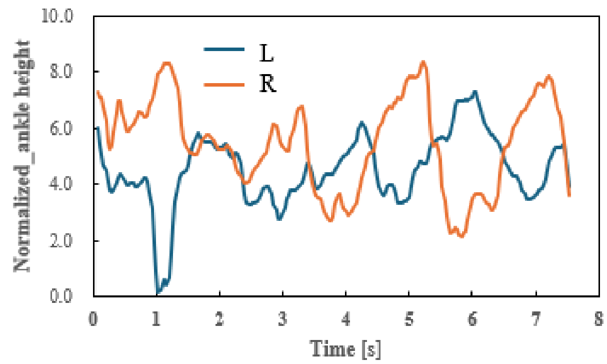


Figure 3. Normalized ankle height during gait measured 109 d after onset.

B. *Normalized shoulder angle*

The results at the beginning of rehabilitation and after five weeks are presented in Figures 4 and 5, respectively. They were obtained from the inner product of vectors using the coordinates of the left shoulder as the origin and the right shoulder angle with respect to the horizontal direction, normalized by the hip width. Difference in blurring was observed at the beginning of walking between the measurements obtained 72 and 109 d after the onset.

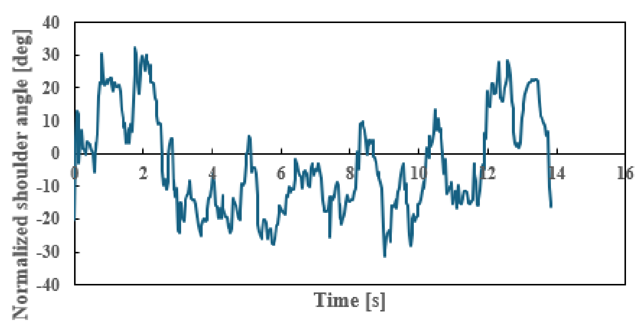


Figure 4. Angle of right shoulder with respect to left shoulder 72 d after onset.

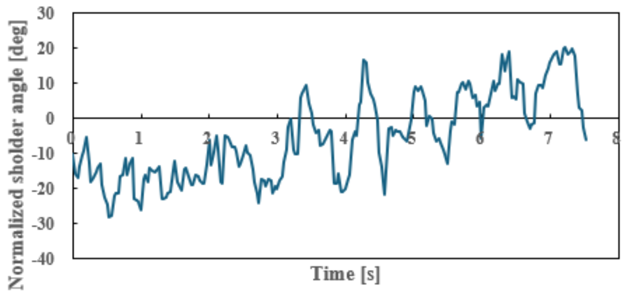


Figure 5. Angle of right shoulder with respect to left shoulder 109 d after onset.

In Figures 4 and 5, the horizontal axis represents time in seconds, and the vertical axis represents the normalized shoulder angle in degrees.

C. *Blurring of nose position*

We observed that the camera position was slightly off the frontal direction in the direction of gait owing to restrictions at the rehabilitation site during filming. Hence, a normalized nose center position was obtained as a linear function, and the difference from the center position was defined as the nose shake.

Figures 6 and 7 show the lateral swing of Participant A's nose during walking, as observed from the frontal direction at 72 and 109 d after onset, respectively. The initial measurement at 72 d after the onset showed minor blurring at the beginning of walking because the participant required more time to begin walking, as compared with the measurement at 109 d after the onset. However, the overall variation in amplitude was approximately similar when viewed over the entire time period. The red and green circles indicate misalignments to the left and right sides of the body, respectively. In Figures 6–11, nose blurring is shown as time on the horizontal axis in seconds, and normalized nose position blurring variance is shown on the vertical axis.

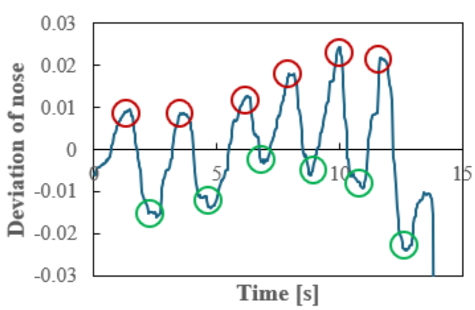


Figure 6. Nasal blurring in left and right directions 72 d after onset.

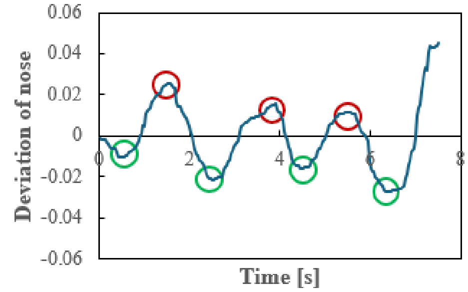


Figure 7. Nasal blurring in left and right directions 109 d after onset.

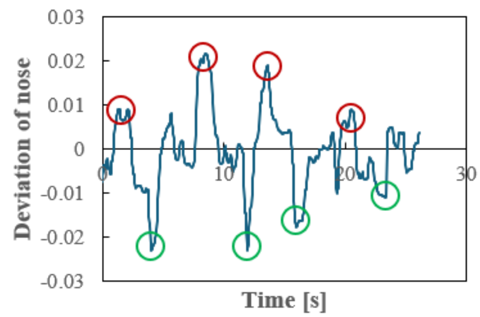


Figure 8. Nasal blurring in left and right directions 46 d after onset.

Figures 8 and 9 show the lateral swing of Participant B's nose during walking, as observed from the frontal direction at 46 and 79 days after onset, respectively.

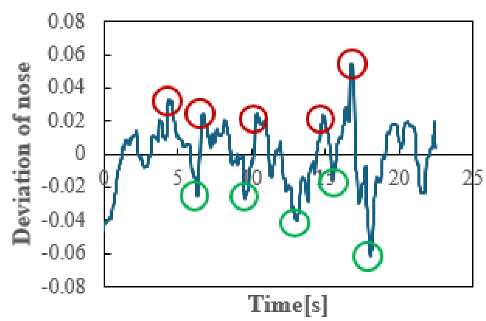


Figure 9. Nasal blurring in left and right directions 79 d after symptom onset.

Regarding the blurring of the nose, the camera was set almost in front of the participant; however, the participant shifted from the center of the screen while walking toward the camera. Figures 6–11 show the first and second measurement results for Participants A, B, and C after calculating the deviation from the approximate straight line and performing corrections, respectively. The red and blue circles in the figure represent the peak of the participant's nasal blurring to the left and right, respectively.

Figures 10 and 11 show the lateral swing of Participant C's nose during walking, as observed in the frontal plane at 44 and 72 d after onset, respectively.

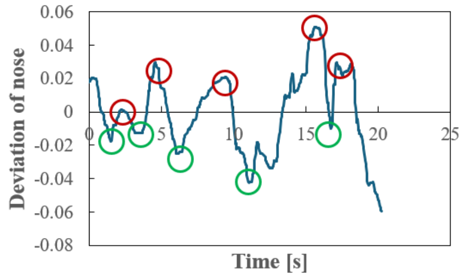


Figure 10. Nasal blurring in left and right directions 44 d after onset.

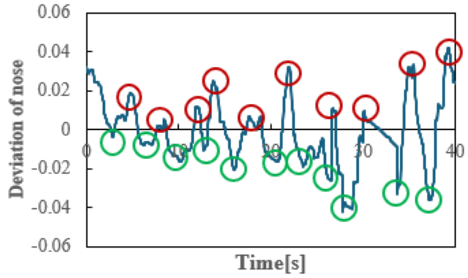


Figure 11. Nasal blurring in left and right directions 72 d after onset.

Table 1. Standard deviation of nasal blurring for each participant.

	1st measurement	2nd measurement
A	1.03E-2	1.65E-2
B	8.74E-3	1.90E-2
C	2.37E-2	1.72E-2

By adopting a methodology reported in the literature, the standard deviation of nasal blurring for each participant during the first and second measurements is as shown in Table 1.

Table 2 shows the average values of the leftward and rightward nasal swings as well as the difference between them for the participants, as indicated by red circles in the graph of the first measurement results.

Table 3 shows the average values of the leftward and rightward nasal swings as well as the difference between them for the participants indicated by red circles in the graph of the second measurement results.

Table 2. Left and right shift results of first measurement.

	Left shift	Right shift	L-R
A	0.0159	-0.0121	0.028
B	0.0109	-0.0188	0.030
C	0.0264	-0.0216	0.048

Table 3. Left and right shift results of second measurement.

	Left shift	Right shift	L-R
A	0.0178	-0.0188	0.037
B	0.0319	-0.034	0.066
C	0.0197	-0.0214	0.041

D. *Change in width between toe and heel*

Figures 12 and 13 show the changes in the width between the toes and heel of the left and right feet at 72 and 109 days after onset, respectively. As depicted, the width was smaller at 109 d after the onset. The horizontal axes in Figures 12 and 13 show the time in seconds, and the vertical axes show the normalized left and right toe widths normalized by the waist width.

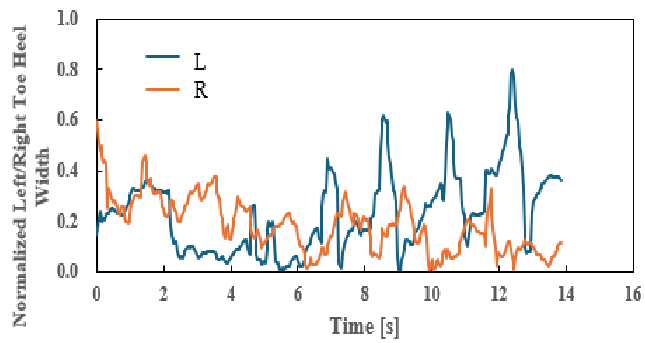


Figure 12. Normalized width between toes and heels of right and left feet 72 d after onset.

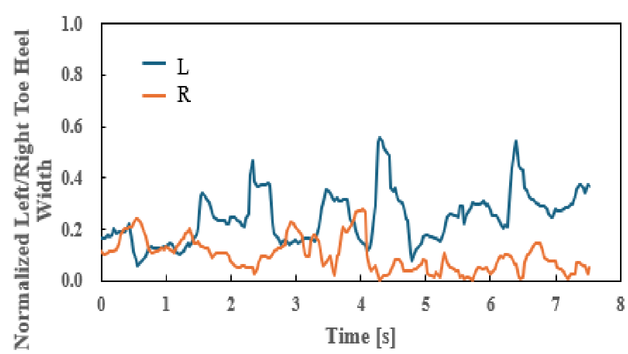


Figure 13. Normalized width between toes and heels of right and left feet 109 d after onset.

E. *Blurring between left and right elbow widths*

Figures 14 and 15 show the blurring between the left and right elbow widths 72 and 109 days after onset, respectively. As depicted, the width decreased at 109 d after the onset. The horizontal axis in Figures 14 and 15 shows the time in seconds, and the vertical axis shows the left and right elbow widths normalized by the waist width.

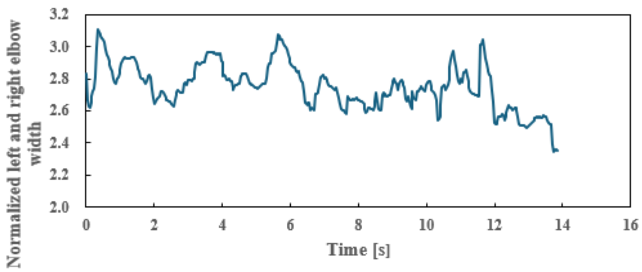


Figure 14. Normalized left and right elbow widths 72 d after onset.

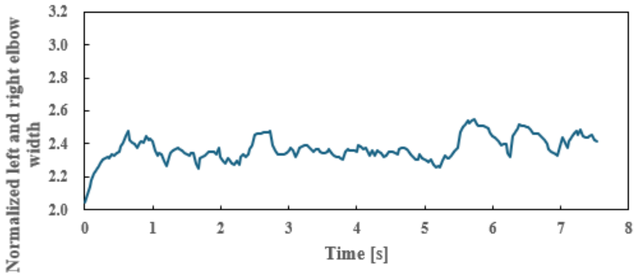


Figure 15. Normalized left and right elbow widths 109 d after onset.

Figures 16 and 17 show the blurring between the left and right elbow widths normalized by the shoulder width 72 and 109 d after onset, respectively. As depicted, the width decreased at 109 d after the onset. The horizontal axis in Figures 16 and 17 shows the time in seconds, and the vertical axis shows the shoulder width normalized by the widths of the left and right elbows.

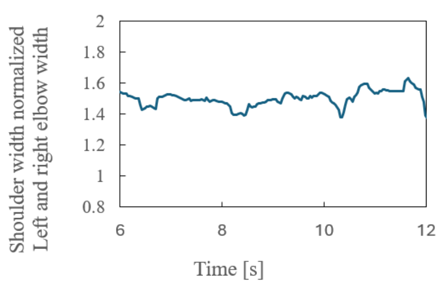


Figure 16. Normalized left and right elbow widths 72 d after onset.

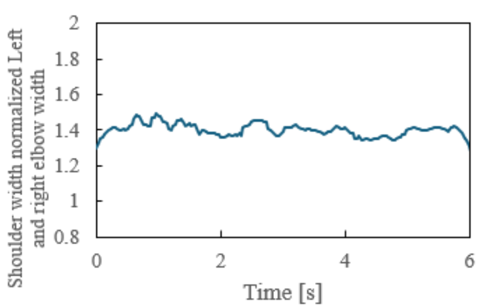


Figure 17. Shoulder-normalized left and right elbow widths 109 d after onset.

F. *Temporal changes in knee and ankle during gait*

Figures 18 and 19 show the left and right knee and ankle heights at 72 and 109 days after onset, respectively, normalized by the waist width. In Figures 18–20, the horizontal axis shows the time, in seconds, and the vertical axis shows the knee and ankle from the normalized waist.

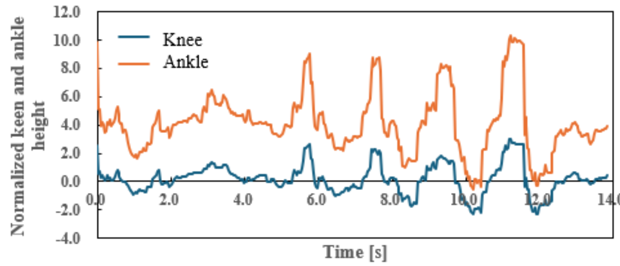


Figure 18. Normalized left knee and left ankle heights 72 d after onset.

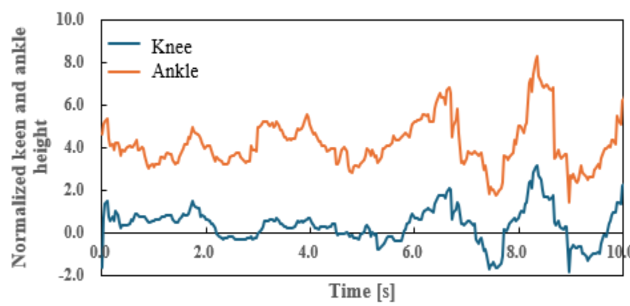


Figure 19. Normalized right knee and right ankle heights109 d after onset.

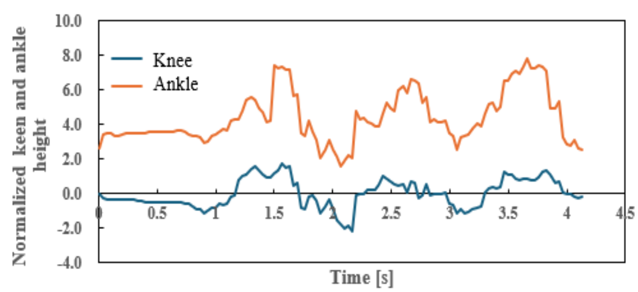


Figure 20. Left knee and ankle heights of Participant D.

Figure 20 shows the left knee and ankle heights normalized by the width of Participant D's waist as a reference.

G. *Example of stride length*

Figures 21 and 22 show the left and right heel heights of Participant B during walking, respectively, normalized by the hip width. They were measured 18 d after symptom onset. In these figures, the inclination was corrected based on distance. Stride length is defined as the distance from the minimum value to the next minimum value. The area indicated by a square in the figure represents the stride length. At approximately 5, 12, and 17 s, for the left ankle shown in the figure, the maximum value was followed by the minimum value, and data that could not be considered as one step were obtained. However, by comparing the data for the right ankle and watching the video, we confirmed that the ankle was not stepping forward owing to greater unsteadiness. Therefore, a wider width was not considered as one step. The horizontal axis in Figures 21–24 shows the time in seconds, and the vertical axis shows the relative position of the ankle with correction.

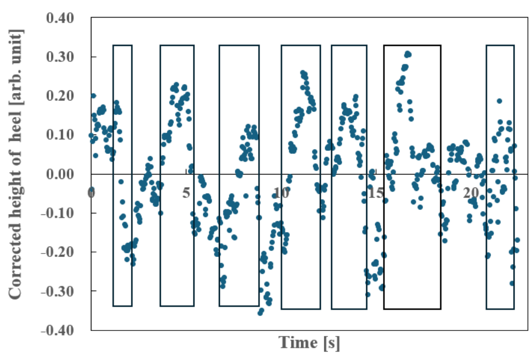


Figure 21. Left-heel height during walking of Participant B normalized by hip width.

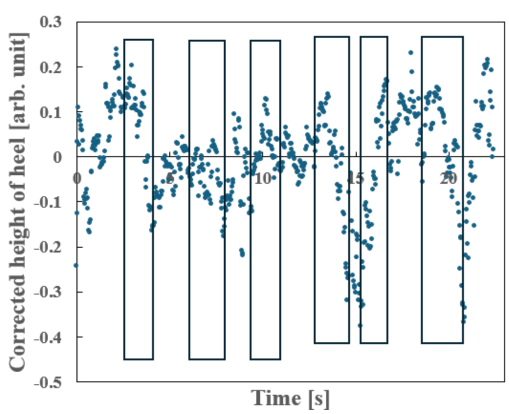


Figure 22. Right-heel height during walking of Participant B normalized by hip width.

Figures 23 and 24 show the left and right heel heights of Participant C during walking, respectively, normalized by the hip width. They were measured 18 d after symptom onset. In these figures, the inclination was similarly corrected based on the distance. Stride length is defined as the distance from the minimum value to the next minimum value. The area indicated by a square in the figure represents the stride length. At approximately 5 s, for the left ankle shown in the figure, the maximum value was followed by the minimum value, and data that could be considered as one step were obtained. However, by comparing the data for the right ankle and watching the video, we confirmed that the ankle was not stepping forward owing to greater unsteadiness. Therefore, a wider width was considered as one step.

Additionally, we viewed the video to determine whether the data for the right foot at 11 and 15–17 s varied. We discovered that the ankle was unsteady and thus could not to identify a clear step; therefore, we excluded it from the calculations of stride length and speed.

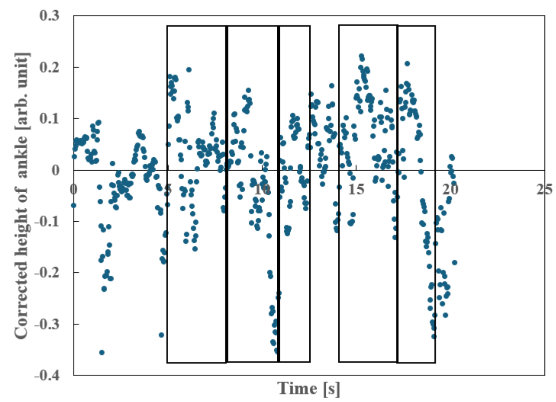


Figure 23. Left-heel height during walking of Participant C normalized by hip width.

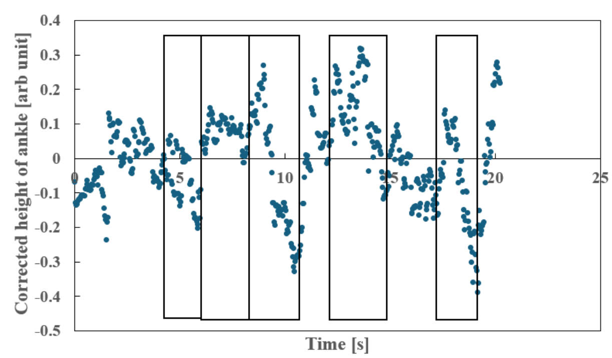


Figure 24. Right-heel height of Participant C during walking normalized by hip width.

The first and second measurements of stride length and foot speed for the three participants (A–C) are listed in Tables 4 and 5. These ratios are listed in Table 6.

Table 4. Step length and speed results for first measurement.

	Left stride length [cm]	Right stride length [cm]	Left foot speed [cm/s]	Right foot speed [cm/s]
Participant A	55.2	83.2	31.8	49.5
Participant B	36.1	35.8	8.7	8.5
Participant C	36.1	35.3	8.8	8.5

Table 5. Step length and speed results for second measurement.

	Left stride length [cm]	Right stride length [cm]	Left foot speed [cm/s]	Right foot speed [cm/s]
Participant A	116.0	83.3	78.3	49.5
Participant B	27.9	28.7	16.3	15.1
Participant C	15.5	16.4	6.5	6.4

Table 6. Ratio of first and second measurements.

	Left stride length ratio	Right stride length ratio	Left foot speed ratio	Right foot speed ratio
Participant A	2.1	1.0	2.5	1.0
Participant B	0.8	0.8	1.9	1.8
Participant C	1.0	0.9	1.0	0.7

H. Smart-insole pressure sensor

For Participant B, we obtained the insole-sensor data during the second walk; Figures 25 and 26 show the results for the left and right foot, respectively. The horizontal axis in Figures 25 and 26 shows the time in seconds, and the vertical axis shows the relative value output from the insole sensor.

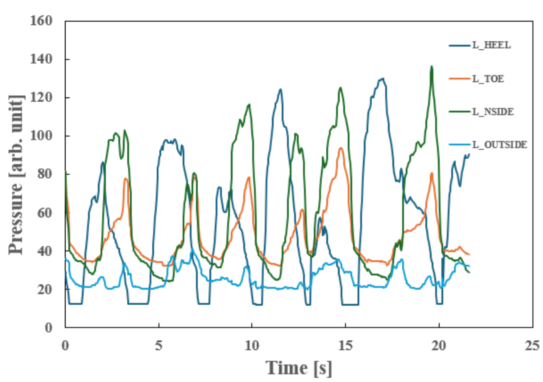


Figure 25. Pressure on four parts of left foot.

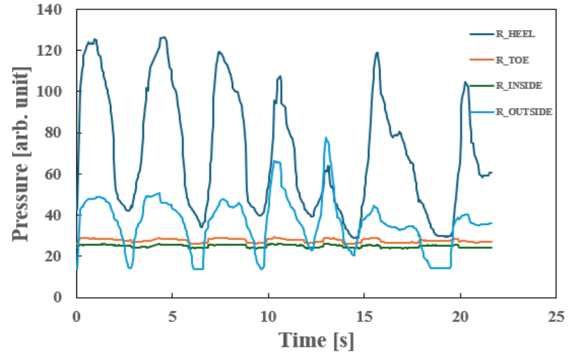


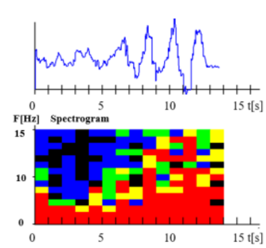
Figure 26. Pressure on four parts of right foot.

The pressure on the outside of the left foot was low. Pressure was applied sequentially from the heel to the inside of the foot and to the toes. However, for the right foot, a high pressure was applied simultaneously to the heel and outside. The pressures on the toes and inside the foot were low, although they were applied at almost the same time.

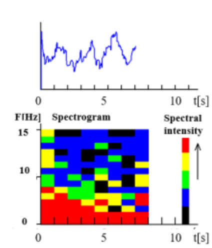
I. Spectrogram

In Figures 27–29, the horizontal axis represents time in seconds, and the vertical axis represents frequency, with different colors indicating the output intensity. The spectral variation over time was calculated via the discrete Fourier transform at every second. The horizontal axis represents the time; the vertical axis represents the frequency; and the spectral intensity (square root of the power spectrum) is shown in red, yellow, green, blue, and black. A rectangular window is used as the time-window function.

The spectrograms of the left and right ankles of Participant A are shown in Figure 27. Owing to the improved foot speed, the second measurement spanned the same distance as the first measurement in approximately 7 s; therefore, the display time was shorter. In Figure 27, (a1) and (a2) show the first and second measurement results for the left ankle, respectively; and (b1) and (b2) show the first and second measurements for the right ankle, respectively.



(a1)



(a2)

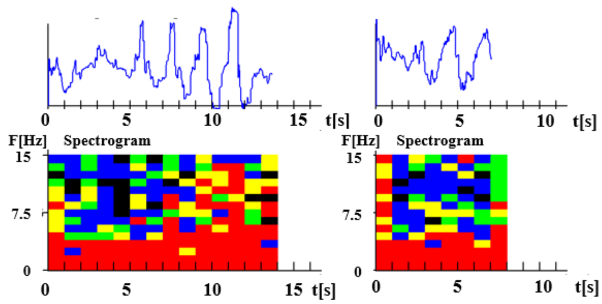
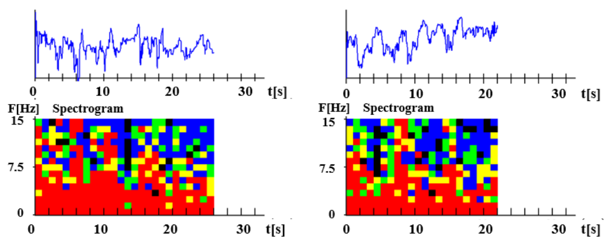
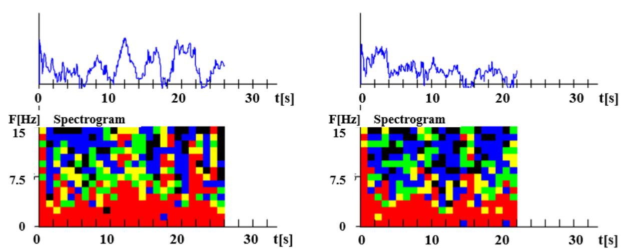


Figure 27. Spectrograms of left and right ankles of Participant A.



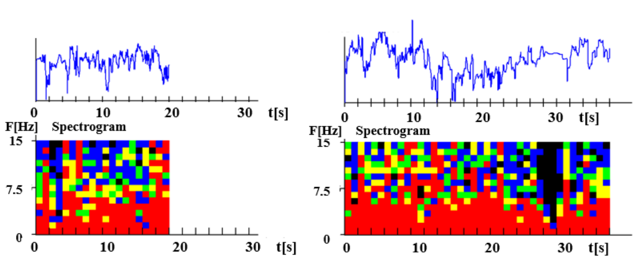
(a1) (a2)



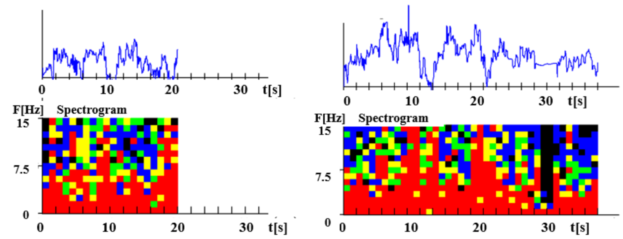
(b1) (b2)

Figure 28. Spectrograms of left and right ankles of Participant B.

The spectrograms of the left and right ankles of Participant C are shown in Figure 29. The second measurement for Participant C was performed approximately one month after rehabilitation. The video showed that compared with the other participants, Participant C received less assistance from the caregiver and was walking carefully while observing his feet, which reduced his foot speed. This resulted in a longer measurement time.



(a1) (a2)



(b1) (b2)

Figure 29. Spectrograms of left and right ankles of Participant C.

IV. DISCUSSION

The authors acknowledge that this study included only three rehabilitated participants (A–C) and one healthy participant. Thus, a cohort study with more participants may be necessary for a more detailed analysis.

We will separately examine the extent to which the differences can be detected between the first and second measurements of the participants using MediaPipe's skeletal recognition data, which were obtained from videos captured from the anterior frontal direction.

A. Difference in normalized left and right ankle heights due to hip width

Normalization by the hip width allowed us to obtain the amplitude of the foot height (even in the early phase) away from the camera, based on which a flat area was observed. This information can be used to determine the stance and swing phases, although a detailed study has not yet been conducted.

B. Normalized shoulder angle

The first-order component was not in the exact frontal direction at the time of the video recording. This might be because the gait began slightly to the left of the center of the screen and eventually shifted to the right. The linear component can be attributed to the gait commencing slightly to the left of the center of the screen and eventually shifting to the right.

C. Blurring of nose position

A previous study showed that using an attachment with an assist function improved leg swing and reduced nose wobble. Therefore, we used the data in the two tables 1 and 2 to examine whether this idea is applicable for improving rehabilitation.

When Participant A's variance of the horizontal value in the travel direction was used to blur the nose position, the blurring width increased slightly, and the variance values were 1.6×10^{-4} and 2.7×10^{-4}. However, based on the data in Tables 3 and 5, we can conclude that the shift in weight from left to right became more balanced and the foot speed improved, owing to an increase in the foot speed 109 d after symptom onset. Therefore, by comparing the left and right deviations of the nose, we can determine whether the weight can be shifted sideways. Considering this value and the deviation in the ankle height, we can evaluate whether the participant was walking in a well-balanced manner.

Considering these two points, which show that Participant A's relative nose position deviation was 1.3 times larger after 109 d and that the speed of the left leg swing increased, as shown in Table 5, we can conclude that Participant A was now able to shift her weight more smoothly. As shown by the blurring of the nose position in Figures 6 and 7, the shift became a large left-right shift smoothly, which is an indicator of recovery. For Participants B and C, the left-right shift was less smooth than that for Participant A. However, this is regarded as an indicator of recovery in conjunction with the content in subsection G, which discusses the stride length and speed.

D. Change in width between toe and heel

In the early phase of rehabilitation, the toe and heel widths of the paralyzed left foot were large, which affected the widths of the right toe and heel. Additionally, 109 d after symptom onset, the left toes and heels became narrower and improved, which caused the right toes and heels to become smaller.

E. Blurring between left and right elbow widths

The elbow width during walking was reduced by approximately 20% between the pre-rehabilitation period and 109 d after symptom onset, thus confirming the beneficial effects of rehabilitation. Based on Figures 16 and 17, normalization by shoulder width was effective, as evidenced by the reduced blur width.

F. Temporal changes in knee and ankle during gait

The temporal difference between the knee and ankle was not evident in this experiment using MediaPipe, although the knee was slightly ahead of the ankle when evaluated on the time axis in some cases. By contrast, the analysis of the experiment conducted with Participant B did not reveal a time difference between the knee and ankle onset of movement.

G. Stride and gait speed

The following were concluded by comparing the left and right stride lengths and left and right foot-movement speeds between the first and second measurements of three participants (A, B, and C): Participant A, who showed recovery, had almost the same stride length and movement speed as the right foot, whereas the stride length and movement speed of the paralyzed left foot showed significant improvement.

Participant B's stride length changed marginally; however, his foot-movement speed increased, thus suggesting that the rehabilitation was effective.

No significant differences were observed in the calculation results for Participant C. Only slight difference was observed, which is attributable to differences in the degree of improvement among Participants A, B, and C and may reflect the degree of recovery.

Comparing the ratios of standard deviations, the standard deviations of the stride length and left-foot speed of Participants A and C decreased from 0.8 to 0.5 in the second measurement as compared with the first, which is considered an indicator of recovery. For Participant B, the ratio of standard deviations for both the stride length and gait speed increased by two to three times. This result, in combination with the insole-sensor results shown in Figures 26 and 27, indicates that the right leg shifted with the non-paralyzed left leg as the center of gravity, thus indicating the importance of using the results in combination with other assessments to evaluate recovery.

H. Smart-insole pressure sensors

Because the participant had right-sided paralysis, the left foot was subjected to pressure from the heel to the outer part and then to the toes. When the participant placed weight on the right foot, pressure was applied almost simultaneously, resulting in the flat-foot condition.

I. Spectrogram

The spectrogram of Participant A showed a significant improvement in the second measurement, although a fewer high-frequency components and a more stable walking pattern were observed.

In the second measurement shown in the spectrogram of Participant B, the paralyzed right leg showed a slight decrease in the number of high-frequency components. For the right leg, the number of high-frequency components decreased during walking, thus suggesting stable movement without blurring. This might reflect the difference in the manner by which weight was applied to the right foot, as shown by the results for the smart insole.

In the spectrogram of Participant C, even in the second measurement, the paralyzed right leg indicated a slight decrease in the number of high-frequency, thus suggesting the effectiveness of rehabilitation.

V. CONCLUSIONS

The effect of rehabilitation was verified using images captured from the frontal direction using MediaPipe. Standardization by hip width reduced the effect of the distance from the camera, and the change in gait was confirmed by determining the ankle height during walking, the tilt of the left and right shoulders, the width between the toes and heels, and the width between the left and right elbows. An evaluation method was demonstrated for cases in which images could not be captured easily from the sagittal plane and could only be captured from the frontal direction owing to limitations in imaging direction, which occurs in actual rehabilitation settings. Additionally, the effectiveness of shoulder normalization was demonstrated.

Walking ability and walking patterns are important for clinical gait assessment. Quantitative instrumental gait analysis is recommended for evaluating gait performance and quality in clinical settings; however, its current application remains insufficient. The results obtained in the current experiments were based on a short rehabilitation period and a small number of participants. To obtain more accurate conclusions, future studies should include a larger sample size and a long-term analysis.

Moreover, we cannot conclude that this assessment method is highly relevant for evaluating walking ability in real-life situations. Nevertheless, one patient demonstrated improvement in real-life walking ability.

ACKNOWLEDGMENT

This study was supported by the JSPS KAKENHI (grant numbers JP20K11924 and JP23K11207). We thank the patients and staff of the Isogo Central Hospital for their cooperation in this study.

REFERENCES

[1] Y. Uchida, T. Funayama, E. Ohkubo, D. Souma, and Y. Kogure, “Detecting Gait Changes with Front-Facing Video and MediaPipe: A Hemiplegic Patient Case Study,” GLOBAL HEALTH 2024, pp. 10-15, IARIA, ISBN: 978-1-68558-189-3.

[2] A. A. Huueck, D. M. Mohan, N. Abdallah, M. E. Rich, and K. Khalaf, “Present and future of gait assessment in clinical practice: Towards the application of novel trends and technologies,” Frontiers in Medical Technology, DOI 10.3389/fmedt.2022.901331, pp. 1-21, 2022.

[3] S. Campagnini et al, “Machine learning methods for functional recovery prediction and prognosis in post-stroke rehabilitation:a systematic review,” Journal of NeuroEngineering and Rehabilitation,” https://doi.org/10.1186/s12984-022-01032-4, pp. 1-22, 2022.

[4] V. Varma and M. Trkov, “Investigation of intersegmental coordination patterns in human walking,” Gait & Posture, vol. 112, pp. 88-94, 2024.

[5] D. Xu et al., “A new method proposed for realizing human gait pattern recognition: Inspiration for the application of sports and clinical gait analysis,” Gait & Posture, vol. 107, pp. 293-305, 2024.

[6] M. Windolf, N. Gotzen, and M. Morlock, “Systematic Accuracy and Precision analysis of Video motion Capturing Systems-Exemplified on The Vicon-460 system,” Journal of Biomechanics, vol. 41, pp. 2776-2780, 2008.

[7] T. B. Rodrigues, D. P. Salgado, C. O. Cathain, N. O’Connor, and N. Murray, “Human Gait Assessment Using a 3D Marker-less Multimodal Motion Capture System,” Multimedia Tools and Applications vol. 79, pp. 2629-2651, 2020.

[8] P. Plantard, E. Auvinet, A. S. Le Pierres, and F. Multon,”Pose Estimation with a Kinect for Ergonomic Stuidies: Evaluation of the Accuracy Using a Virtual Mannequin,” Sensors, pp. 1785-1803, 2015.

[9] R. A. Clark, B. F. Mentiplay, E. Hough, and Y. H. Pus, “Three-Dimensional Cameras and Skeleton Pose Tracking for Physical Function Assessment: A Review of Use, Validity, Current Developments and Kinect Alternatives,” Gait & Postture, vol. 68, pp. 193-200, 2019.

[10] J. A. Albert el al., “Evaluation of the Pose Tracking Performance of the Azure Kinect and Kinect v2 for Gait Analysis in Comparison with a Gold Standerd: A Pilot Study, Sensors, vol. 20, 5104, 2020.

[11] Y. Ma, K. Mithratatne, N. Wilson, Y. Zhang, and X. Wang, “Kinect v2-Based Gait Analysis for Children with Cerebral Palsy: Validity and Reliability of Spaial Margin of Stability ad Spationtemporal Vaiables,” Sensors, vol. 21, 2104, 2021.

[12] D. Imoto, S. Hirano, M. Mukaino, E. Saitoh, and Y. Otaka, “A Novel Gait Analysis System for Detecting Abnormal Hemiparetic Gait Patterns during Robo-assisted Gait Training : A Criterion Validity Study among Healthy Adults,” Frontiers in Neurorobotics, 16:1047376, 2022.

[13] G. Hidalgo et al., “CMU-Perceptual-Computing-Lab,” Avairable: https://gitgub.com/CMU-Preceptual-Computing-Lab/openpose, retrieved: August, 2024.

[14] Z. Cao, G. Hidalgo, T. Simon, S-E. Wei, and Y. Sheikh, “OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields,” IEEE Transaction on Patern Analysis and Machine Intelligence, vol. 43, pp. 172-186, 2021.

[15] N. Nakano et al., “Evaluation of #D Markerless Motion Capture Accuracy Using OpenPose With Multiple Video Cameras,” Frontiers in Sport and Active Living, vol. 20, Article 50, 2020, doi:10.3389/fspor.2020.00050.

[16] M. Ota, H. Tateuchi, T. Hahiguti, and N. Ichihasi, “Verification of validity of gait analysis systems during treadmill walking and running using human pose tracking algorithm,” Gait and Posture, vol. 85, pp. 290-297, 2021.

[17] Y. Saiki et al., “Reliability and validity of OpenPose for measuring hip knee ankle angle in patients with knee osteoarthritis,” Scientific Reports, vol. 13, 3297, 2023.

[18] V. Bazarevsky et al., “BlazePose: On-device Real-time Body Pose tracking,” arXiv:2006.1204v1 [cs.CV] 2020.

[19] G. Kaur, G. Jaju, D. Agawal, K. Lyer, and C. M. Prashanth, “Implementation of Geriatric Agility Detection Using MediaPipe Pose,” International Journal of Recent Advances in Multidisciplinary Topics, vol. 3, 119, 2022, ISSN:2582-7839.

[20] Y. Uchida, T. Funayama, and Y. Kogure, “Investigation of the Application of MediaPipe to Gait Analysis,” pp. 1-6, IARIA, 2022. ISBN: 978-1-61208-995-9.

[21] Y. Uchida, T. Funayama, and Y. Kogure, “Possibility of Gait Analysis with MediaPipe and Its Application in Evaluating the Effects of Gait-assist Devices,” pp. 44-54, IARIA, International Journal on Advances in Life Sciences, vol. 15, no. 1 & 2, 2023.

[22] J.-L. Vhung, L.-Y. Ong, and M-C. Leow, “Comparative Analysis of Skelton-Based Human Pose Estimation,” Future Internet, vol. 14, 380, 2022.

[23] Y. Uchida, T. Funayama, E. Ohkubo, and Y. Kogure, “Considerations for Applying MediaPipe to Gait Analysis,” GLOBAL HEALTH 2023, pp. 12-17, IARIA.

[24] S. Kim et al., “Assessing physical abilities of sarcopenia patiens using gait analysis and smart insole for development of digital biomarker,” Scientific Reports, vol. 13, 10602, 2023, https://doi.org/10.1038/s41598-023-37794-7.

[25] Y. Fan, “The Gesture Recognition Improvement of Mediapipe ModelBased on Histrical TrajectoryAssist Tracking Kalman Filtering and Smooth Filterling,” CISAI 2024, Shaoxing, China, pp. 641-674.

[26] M. Lei, Z. Wang, and F. Chen, “Ballet Form Traning Based on MediaPipe Body Posture Monitoring,” Journal of Physics: Conference Series, ICAITA-2023, 2637(2023)012019, doi:10.1088/1742-6596/2637/1/012019.

[27] D. Deng et al., “Interpretable video-based tracking and quantification of parkinsonism clinical motor stated,” npj Parkinson’s Disease, (2024) 10:122, https://doi.org/10.1038/s41531-024-00742-x.

[28] M. U. Friedrich et al., “Validation and application of computer visopn algorithms for video-based tremor analysis,” npj Digital medicine, (2024)7:165, https://doi.org/10.1038/s41746-024-01153-1.

[29] G. Colnella-Barba et al., “Using a Webcam to Assess Upper Extremity Proprioception: Extperrimental Validation and Application to Persons Post Stroke,” Sensors, vol.24, 7434, 2024.

[30] T. Funayama, Y. Uchida, Y. Kogure, D. Souma, and R. Kimura, “Exploring the Assessment of Steps Using Insoles with Four-Part Pressure Sensors,” Sensors & Transducers, vol. 263, pp. 21-28, 2023.

Integrating Multiple Intelligences and I-TRIZ

A Framework for Developing Individualized Problem-Solving Skills in Vocational Training

Norikatsu Fujita
Ability Development Dept
The Polytechnic University of JAPAN
Tokyo, Japan
e-mail: fujita@uitec.ac.jp

Sho Aoki
Electronics and Information Systems Dept
Shikoku Polytecnic College
Kagawa, Japan
e-mail: Aoki.sho@jeed.go.jp

Abstract - **The paper proposes an integrated pedagogical approach to develop individualized problem-solving skills in vocational training, specifically addressing the growing number of trainees with special needs, such as developmental disabilities. The proposed approach integrates a framework of "fundamental skills," derived from Gardner's Multiple Intelligences (MI) theory, with I-TRIZ, a structured problem-solving methodology. Using interviews with vocational trainers and the Steps for Coding and Theorization (SCAT) method, we redefined the "fundamental skills" as 26 practical skills, which were then categorized under six intelligences from MI theory. This paper illustrates the framework's potential through a hypothetical case study in welding training, demonstrating how trainees with disabilities can strategically apply these skills at each stage of the I-TRIZ process to acquire effective and autonomous problem-solving abilities. Ultimately, this integrated framework offers a promising pathway for developing individualized support and fostering self-regulated problem-solving skills in vocational training.**

Keywords-Multiple Intelligences(MI)Theory; Ideation-TRIZ; Developmental disabilities; Self-Regulated Learning; Polytechnic science.

I. Introduction

Practical problem-solving skills, such as addressing workplace challenges and adapting flexibly to change, are essential for sustained career success in a world of rapid technological innovation and globalization [2].

In vocational training, the need for individualized instruction is growing due to the increasing number of trainees requiring special support—including those with developmental disabilities—and the diversification of trainee backgrounds. This paper is a substantially extended version of research originally presented at an international conference [1]. Trainees exhibit diverse attributes—such as vocational history, educational background, and age—as well as significant variations in learning styles, prior knowledge, and skill levels. Therefore, it is crucial to cultivate general-purpose problem-solving skills that leverage individual strengths and prepare trainees to manage unexpected situations, ensuring all can achieve independence and success.

Conventional instruction in vocational training has faced three major challenges [3]. First, instruction often relies on heuristic, case-specific methods rather than established systematic approaches. Second, it is difficult to tailor instruction to the unique characteristics and learning needs of each trainee. Third, excessive support from instructors may inadvertently hinder the development of trainee independence. Therefore, even if trainees overcome challenges during training, they may fail to develop the ability to think and solve problems independently, leading to subsequent difficulties after employment. While indicators for assessing individual characteristics exist in other fields [4], there is a notable scarcity of such indicators developed from a pedagogical perspective suitable for adaptation to vocational training environments.

Gardner's multiple intelligences (MI) theory [5], which views human intelligence from multiple perspectives, has potential for addressing this situation. However, the abstract nature of MI theory makes it difficult to translate into concrete instructional practices or to quantitatively measure its educational effectiveness in vocational training contexts [6].

To address these challenges, this study proposes a pedagogical approach that integrates MI theory with the I-TRIZ methodology to foster both individualized instruction and general-purpose problem-solving skills. This integrated approach is designed to directly counteract the identified challenges. Specifically, the MI-based framework provides a systematic basis for tailored instruction, addressing Challenge #2. Simultaneously, the I-TRIZ process establishes the systematic, non-heuristic methodology that conventional instruction lacks, thereby addressing Challenges #1 and #3. This approach is centered on the concept of "fundamental skills," which are derived from MI theory to address the unique strengths and challenges of individual trainees. By using I-TRIZ [7], a structured problem-solving methodology, it becomes possible to better support trainees in mastering the process of identifying root causes and devising concrete solutions. The "fundamental skills" were defined through a detailed analysis of 192 problematic behaviors observed in vocational training settings [8], allowing us to translate the abstract nature of MI theory into concrete, applicable instructional components.

The proposed approach builds on a history of pedagogical development at Polytechnic University, which has offered training focused on fundamental skills since 1998. This long-standing program has consistently integrated methods aligned with MI theory and was updated in 2022 to refine its framework from 32 to 26 core skills, addressing issues of redundancy and granularity. This continuous adaptation underscores the sustained significance of this pedagogical model in vocational education.

This paper is structured as follows. Section II outlines the theoretical foundations of the proposed approach, integrating MI theory and I-TRIZ. Section III details the methodology used to redefine the fundamental skills. Section IV presents a hypothetical case study to illustrate the practical application of the framework. Section V provides a broader discussion, positioning the contribution within the context of prior and related work. Section VI acknowledges the study's limitations and suggests directions for future research. Finally, Section VII offers concluding remarks.

II. Theoretical Framework and Proposed Approach

This section outlines the theoretical underpinnings of the proposed pedagogical approach, drawing upon Gardner's Multiple Intelligences (MI) theory and the I-TRIZ problem-solving methodology. It defines "fundamental skills" within the context of MI theory and explains how these skills are integrated with I-TRIZ to facilitate individualized learning and problem-solving in vocational training.

A. MI Theory and Fundamental Skills

Gardner’s MI theory posits that human intelligence is not a single, uniform ability, but rather a collection of independent intelligences. Gardner initially proposed seven intelligences (linguistic, logical-mathematical, spatial, bodily-kinesthetic, musical, interpersonal, and intrapersonal), and later added naturalist intelligence. Other candidate intelligences, such as existential intelligence, have also been considered by Gardner [9]. These intelligences reflect humans’ diverse approaches to processing information and solving problems.

Within Gardner's Multiple Intelligences (MI) theory, an "intelligence" refers to an innate biopsychological potential for an individual to solve problems or create products that are valued within a particular cultural setting. In contrast, a "skill" in this study is defined as the observable mastery of a specific behavior, which emerges from this latent potential and is developed through training and experience. In other words, a skill is a tangible manifestation of an intelligence, measurable and evaluable in a practical context.

This study focuses on six intelligences from the MI theory (linguistic, logical-mathematical, spatial, bodily-kinesthetic, interpersonal, and intrapersonal) that are most relevant to vocational manufacturing training. In our framework, we operationalize these intelligences through a set of “fundamental skills,” which are concrete, observable behaviors that manifest from an individual's latent potential. Rather than treating each intelligence as a broad, overarching concept, we defined four to five specific “fundamental skills” corresponding to each of the six intelligences. For example, “understanding instructions (using demonstratives)” is a fundamental skill related to linguistic intelligence, while “identifying 2D shapes” is a skill related to spatial intelligence. These fundamental skills are measurable behavioral indicators. Defining them allows for a more detailed understanding of each trainee’s abilities and enables more practical, individualized instruction. This approach could be particularly beneficial for learners with developmental disabilities, who often have uneven profiles of cognitive strengths and weaknesses [10]. This approach can teach them to leverage their strengths to compensate for their weaknesses. It also promotes self-regulated learning by enabling trainees to understand their own learning styles and choose strategies that best suit their needs.

This principle is exemplified by a case study [11] of an individual with developmental disabilities who had a deficient body image. This condition hindered his coordinated movement. For this individual, logical-mathematical instruction (e.g., verbalizing the cause-and-effect of movements) was used to overcome challenges related to bodily-kinesthetic intelligence. This approach enabled the individual to understand movement mechanics and ultimately master complex athletic skills, demonstrating that learners can leverage strengths in one intelligence to compensate for difficulties in another.

B. Utilizing I-TRIZ and Fundamental Skills for Problem-Solving

While MI theory provides the crucial lens for understanding individual cognitive differences ('what' to focus on), it does not offer a structured process for problem-solving. Conversely, while I-TRIZ provides a systematic process ('how' to solve), it does not inherently account for individual learner characteristics. Therefore, the synergy of these two theories is essential to our approach, as it creates a comprehensive pedagogical model that integrates individualized diagnosis with systematic problem-solving. The proposed approach uses I-TRIZ, a simplified version of TRIZ, as a structured problem-solving methodology. TRIZ is a theory derived from the analysis of a vast number of patents to extract common patterns of invention. I-TRIZ retains the core principles of TRIZ and consolidates TRIZ’s 40 inventive principles into nine key principles, enabling a simpler, more practical approach to problem-solving. This study adapts a problem-solving process based on I-TRIZ, consisting of the following steps:

1. Define Objectives and Constraints: Clearly define the problem to be solved and understand the goals to be achieved, the current situation, the ideal outcome, and constraints (time, budget, resources, etc.).

2. Model the Problem: Visualize the problem and its related factors using tools such as diagrams, and identify the root cause through methods like repeated questioning. At this stage, fundamental skills related to logical-mathematical intelligence, such as “Subdivision of information,” and those

related to interpersonal intelligence, such as "understanding others' intentions," are important.

3. Generate Ideas: Utilize I-TRIZ's nine inventive principles (segmentation, extraction, local quality, asymmetry, combination, universality, nesting, prior action, and counteraction/inversion). These principles provide structured pathways for creative problem-solving across diverse domains [12]. The strategic application of a trainee's fundamental skills can significantly enhance their ability to use these principles. By mapping specific skills to each inventive principle, instructors can guide trainees to leverage their cognitive strengths.

For example, the principle of Segmentation involves dividing a system or problem into smaller, independent, or more manageable parts. This directly engages Skill 7: Subdivision of information (Logical-mathematical), which is the ability to break down complex information. Additionally, visualizing how a physical object can be taken apart or how distinct modules can be formed relies on Skill 11: Identifying 3D shapes (Visual-spatial). If the segmentation involves physically manipulating components, Skill 14: Hand dexterity (Bodily-kinesthetic) becomes crucial.

The principle of Extraction focuses on identifying and separating a necessary part/property or removing an interfering one. To apply this, a trainee might use Skill 9: Understanding of priorities (Logical-mathematical) to determine which elements are essential or detrimental. If the part to be extracted is visually distinct (e.g., color-coded), Skill 12: Identifying color (Visual-spatial) would be beneficial. Furthermore, deciding to remove a long-standing but problematic feature might require Skill 25: Emotional control (Intrapersonal) to overcome attachment to existing designs.

Consider Prior Action, which means performing a required change or action in advance. This principle encourages foresight and planning. Trainees strong in Skill 23: Understanding future prospects (Intrapersonal) can more readily anticipate future needs or steps, making prior actions intuitive. Executing these preparatory steps often involves understanding and following sequences, supported by Skill 3: Understanding instructions (Linguistic-verbal), and in a workshop context, preparing materials might involve Skill 6: Basic arithmetic operations (Logical-mathematical) for measurements or counting.

The principle of Combination involves merging or assembling different elements or functions. This can be effectively utilized by trainees who excel at Skill 8: Completion of unclear points (Logical-mathematical), as they can see how disparate parts might fit together to form a complete solution. If combining ideas from multiple team members, Skill 21: Understanding of others' intentions (Interpersonal) is key. For a trainee with strong Skill 24: Relating to similar experiences (Intrapersonal), they might combine current ideas with solutions from past, analogous problems.

These examples demonstrate that the effective application of I-TRIZ principles is not a purely abstract exercise but can be significantly enhanced by consciously engaging relevant fundamental skills. By understanding these connections, an instructor can guide a trainee to focus on principles that align with their cognitive strengths. For instance, a trainee strong in Visual-Spatial intelligence might be particularly adept with principles like Nesting or Asymmetry (requiring Skill 10 & 11), while one with strong Logical-mathematical intelligence might find Segmentation or Prior Action more accessible (utilizing Skill 7 & 9). This strategic application transforms the inventive principles from an abstract list into a personalized toolkit for innovation, drawing upon a wide range of intelligences.

4. Develop Solution Plans: Evaluate and select multiple ideas under realistic constraints (cost, time, technology, human resources, etc.) and develop them into an executable plan. The optimal solution is determined by comprehensively considering feasibility, effectiveness, risks, and the trainee's own characteristics and abilities.

5. Evaluate Results: Evaluate the achievement level of the implemented solution from multiple perspectives, including numerical targets, feedback from stakeholders, and self-reflection. Based on the evaluation, identify areas for improvement and review the entire process.

In problem-solving and creative activities, it is crucial to understand which of a trainee's "fundamental skills" to use and how to use them effectively. This requires strong intrapersonal intelligence—the ability to accurately understand the trainee's own emotions, motivations, strengths, weaknesses, and goals, and to use this understanding to guide their actions. However, individuals with certain developmental disabilities often face challenges in developing intrapersonal intelligence [13] [14], which can hinder their ability to engage in effective self-reflection and, consequently, impact their problem-solving capabilities. This focus on self-awareness is critical throughout the problem-solving process. Therefore, the proposed approach provides concrete support for developing these crucial skills through strategies such as promoting self-questioning (e.g., asking "why" repeatedly to oneself), incorporating feedback from others, and utilizing self-assessment checklists. These methods aim to foster objective self-review of actions and behaviors. For trainees with developmental disabilities who may struggle with self-reflection, instructors play a vital role by facilitating this process through guided questioning, structured feedback, and assistance in identifying and articulating their own strengths and weaknesses. It also promotes collaborative learning, in which trainees observe each other's work, exchange opinions, and provide advice, thus deepening their learning from multiple perspectives.

III. Redefining Fundamental Skills

This section details the process of redefining the fundamental skills derived from MI theory, making them more concrete and applicable to vocational training contexts. It outlines the methodology used, including semi-structured interviews with vocational trainers and the application of Steps for Coding and Theorization (SCAT) for qualitative data analysis. The section culminates in the presentation of the 26 redefined fundamental skills, categorized under six

MI intelligences, and explains how these skills were refined from the initial 32 skills.

A. Methodology

We conducted semi-structured interviews, each lasting approximately one hour, with 11 novice vocational trainers. All interviews were recorded and detailed transcripts were created. Novice instructors were chosen for this study because they are more likely to face instructional challenges. Consequently, they are expected to be more acutely aware of the fundamental skills required to address these challenges than their experienced counterparts. For each fundamental skill, interviewees were asked to imagine trainees with extremely high and extremely low levels of that skill and describe their specific behaviors in detail (e.g., during practical exercises, break times, interactions with others).

This method was designed to clarify the specific behavioral range associated with each skill and eliminate overlaps or inconsistencies.

This study used SCAT [15], a qualitative data analysis method, to redefine the 32 fundamental skills identified in previous research [16] to better suit the vocational training context. SCAT is a systematic method for extracting concepts from textual data, clarifying their relationships, and building a theory based on those relationships. For instance, its utility in structuring qualitative analysis to explore complex educational outcomes has been demonstrated by Kosaka and Nakawa [17] in their life story analysis of the long-term effects of a secondary education project in Kenya. Following this systematic approach, we applied the SCAT method, which proceeds along the following four steps:

1) Noteworthy Words: Identifying keywords and phrases related to trainees' difficulties (e.g., Explanation/safety part /misleading) from the interview transcripts.

2) Rephrasing: Converting these words into more general terms used in vocational training (e.g., Colloquialisms /points /not understood).

3) Explanatory Concepts: Assigning explanatory concepts to connect these phrases to the vocational training context (e.g., Practical training situations /safety management / trainee cases).

4) Themes and Concepts: Deriving the theme (constituent concept) of "understanding main points in spoken language" through this analysis.

Figure 1 illustrates the SCAT analysis process for redefining the fundamental skill "understanding key points from conversations and texts". The compositional concepts of the results of SCAT of the 11 interviews about "Understanding key points from spoken and written language" are shown in Figure 2. These elements, identified through our SCAT analysis, were derived from interviews where 11 instructors described the behaviors of trainees with extremely high or low proficiency in this skill. The analysis revealed that these 13 elements—such as understanding oral explanations, comprehending information from context, and misunderstanding colloquialisms—collectively define the scope of this fundamental skill and the concrete behaviors it encompasses.

Notably, the difficulties identified by instructors varied, with some emphasizing the inability to grasp main points while others focused on misunderstanding colloquial language. This variation underscores the need for individualized instruction tailored to each trainee's specific challenges. The overarching goal of this study was to identify a comprehensive set of fundamental skills, including the one detailed here, that could collectively explain the 192 problematic behaviors identified in our previous research. Ultimately, the goal of this iterative process was to create a refined set of skills that was not only comprehensive in its explanatory power but also practical and relevant for vocational training instructors.

B. Redefined Fundamental Skills

Figure 3 illustrates the 26 fundamental skills redefined through the SCAT analysis and instructor interviews, as described in Section III.A. These skills represent a refinement of the 32 skills identified in previous research [15], with modifications based on the specific needs and challenges observed in vocational training settings. These 26 skills are categorized under six of Gardner's Multiple Intelligences (MI): Linguistic-Verbal, Logical-Mathematical, Visual-Spatial, Bodily-Kinesthetic, Interpersonal, and Intrapersonal. This categorization is based on the observed behaviors and cognitive processes associated with each skill, as identified through the qualitative data analysis. The figure provides the MI category, skill name, and examples of specific vocational training situations where each skill is

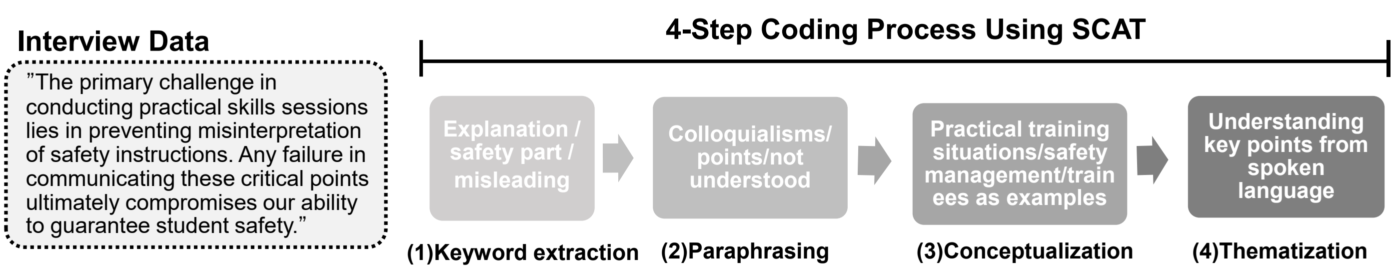


Figure 1. The SCAT analysis process for redefining the fundamental skill "understanding key points from conversations and texts"

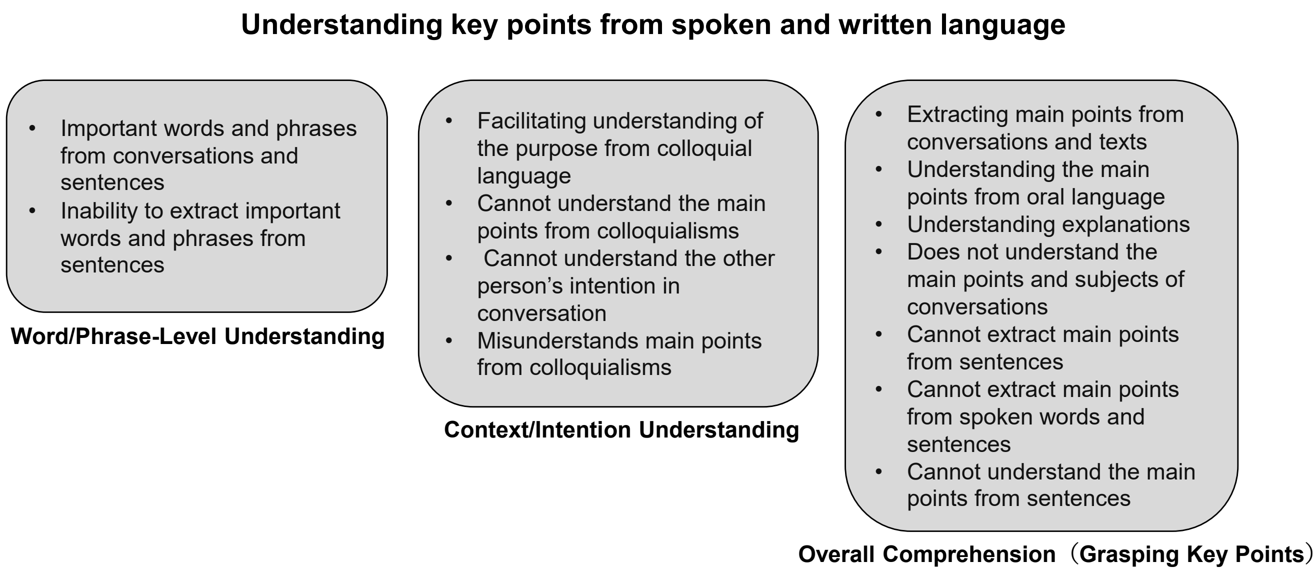


Figure 2. Compositional Concepts of the results of SCAT of the 11 interviews about "Understanding key points from spoken and written language"

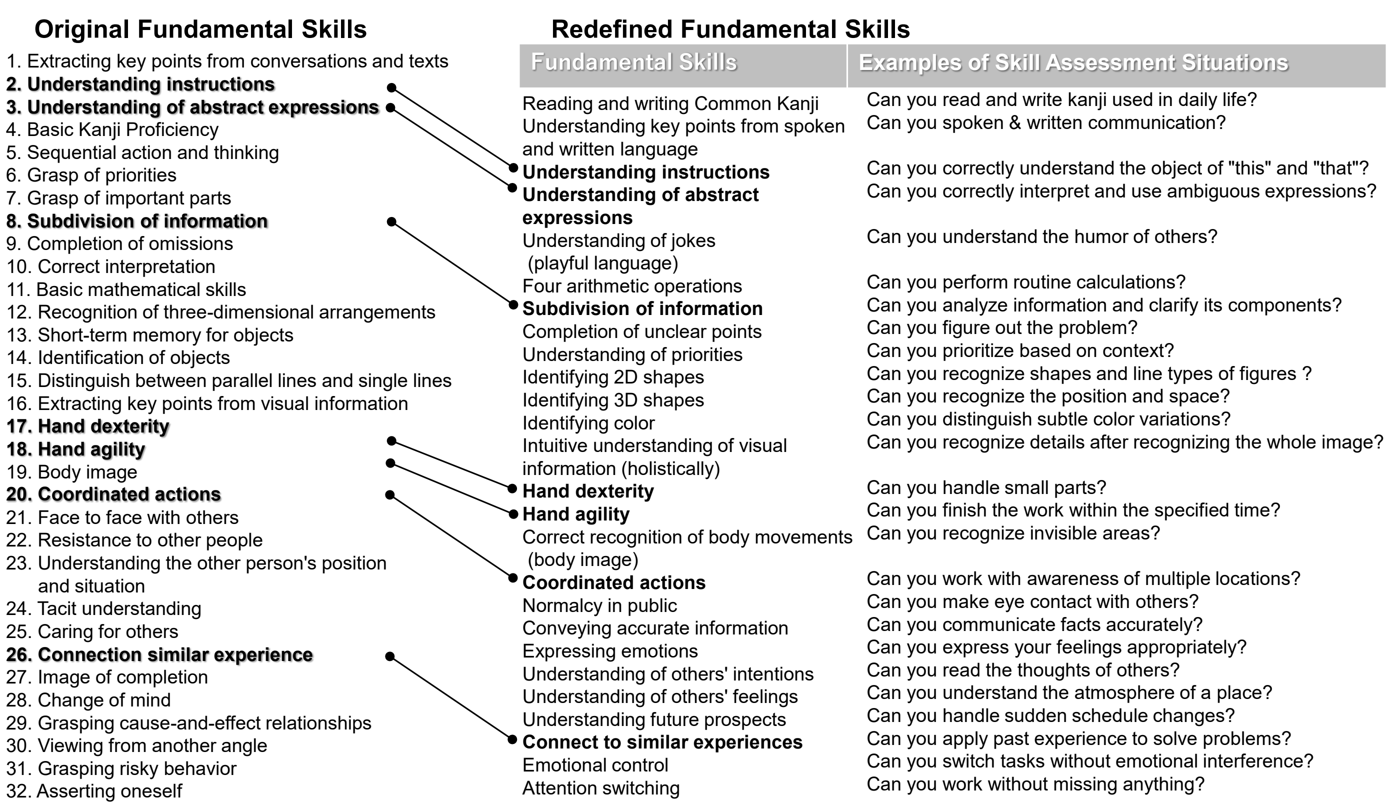


Figure 3. The original and redefined fundamental skills using SCAT analysis

Table 1. Modifications in redefining fundamental skills (changes, integrations, deletion, and additions)

MI	Fundamental Skills	Scope	Reason for improvement	Refined Skills name
Linguistic-verbal	Extracting key points from conversations and texts	・Smoothly understand the purpose from colloquial language. ・Inability to understand the other person's intention in conversation.	**Change:** Focus on any spoken, not just conversational, language.	Understanding key points from spoken and written language
Linguistic-verbal	Basic Kanji Proficiency	・Cannot comprehend Kanji characters at the vocational level. ・Can read and write daily life level Kanji characters.	**Change:**Vocational training level within compulsory education scope	Reading and writing of kanji for common use
Linguistic-verbal			**Addition:** Difficulty conveying jokes in conversation.	Understanding of jokes (playful language)
Logical-mathematical	Sequential action and thinking	・Can work according to the established sequence. ・Deviates from work procedures.	**Integration:** Includes "Understanding of priorities".	Understanding of priorities
Logical-mathematical	Grasp of priorities	・Can determine priorities based on efficiency. ・Cannot determine priorities based on deadlines.	**Integration:** Includes "Understanding of priorities".	Understanding of priorities
Logical-mathematical	Grasp of important parts	・Can grasp the important parts of things. ・Can understand the important parts of multiple tasks.	**Integration:** Includes "Understanding of priorities".	Understanding of priorities
Logical-mathematical	Completion of omissions	・Can understand the lack of processing and improve it. ・Does not have the knowledge to accomplish the objectives.	**Change:** Based on instructor feedback ("unclear points").	Completion of unclear points
Logical-mathematical	Basic mathematical skills	・Can perform four arithmetic operations including fractions. ・Can perform the four arithmetic operations.	**Change:** Based on instructor feedback ("Four arithmetic operations").	Four arithmetic operations
Logical-mathematical	Correct interpretation	・Make sense of numbers by calculation. ・Interpret the significance of things correctly.	**Integration:** Includes "Suddivision of information".	Suddivision of information
Visual-spatial	Recognition of three-dimensional arrangements	・Inability to perceive three-dimensional distance and depth. ・Cannot visualize three-dimensional shapes.	**Change:** Separated into 2D and 3D skills.	Identifying 3D shape
Visual-spatial	Short-term memory of objects	・Cannot memorize numerical values in the short term. ・Can memorize the shape of objects in the short term.	**Delete:**Memory is not included in the MI categories.	
Visual-spatial	Distinguish between parallel lines and single lines	・Understanding of line types is a prerequisite. ・Cannot distinguish shapes.	**Addition:** Separated into 2D and 3D skills.	Identifying 2D shape
Visual-spatial	Distinguish between Objects	・Cannot distinguish colors close to the primary color. ・Cannot distinguish colors.	**Change:** Requires color identification, alongside 2D and 3D skills.	Identifying Color
Visual-spatial	Extraction of key points from visual information	・Cannot focus on hazardous areas. ・Cannot focus on important points.	**Change:** Added holistic, intuitive visual understanding.	Intuitive understanding of visual information (holistically)
Body-kinesthetic	Body image	・Imitates the body movements of others. ・Does not have proper awareness of body range.	**Change:** "Body image" to clearer term.	Correct recognition of body movements (body image)
Interpersonal	Face to face with others	・Inability to use appropriate body language. ・Inability to express appropriate emotions.	**Change:** Based on instructor feedback (a lot of conversation and body language).	Conveying accurate information
Interpersonal	Resistance to other people	・Is uncomfortable in conversation. ・Cannot establish positive rapport with new acquaintances.	**Change:** Focus on attitude	Normalcy in public
Interpersonal	Understanding the other person's position and situation	・Cannot anticipate the feelings of others and act accordingly. ・Cannot understand the other person's situation.	**Change:** Focus on intentions	Understanding of others' intentions
Interpersonal	Tacit understanding		**Change:** Focus on feelings	Understanding of others' feelings
Interpersonal	Caring for others	・Can build good relationships. ・Has difficulty in building good relationships.	**Change:** Emphasizes emotional expression.	Expressing emotions
Intrapersonal	Image of completion	・Clearly defines the target quality. ・Can grasp the process from the image of completion.	**Change:** Supports intuitive visual understanding.	Understanding future prospects
Intrapersonal	Change of mind	・Can switch from break to work quickly. ・Cannot let a bad event drag on.	**Change:** Requires diverse emotional control.	Emotional control
Intrapersonal	Grasp of cause-and-effect relationships	・Cannot investigate the cause of work failure. ・Cannot understand the structure of equipment.	**Integration:** Includes "Connection to similar experiences".	Connection to similar experiences
Intrapersonal	Viewing from another angle	・Flexible in dealing with problems instead of sticking to existing methods. ・Cannot consider multiple means to solve a problem.	**Integration:** Includes "Understanding future prospects".	Understanding future prospects
Intrapersonal	Grasp of risky behavior	・Can predict and avoid hazards. ・Can anticipate danger.	**Change:** Includes "Attention switching".	Attention switching
Intrapersonal	Asserting oneself	・Cannot express physical condition. ・Needs to improve the work environment.	**Integration:** Includes "Normalcy in public".	Normalcy in public(Interpersonal)

utilized and can be evaluated. This figure serves as a key reference point for understanding the individual strengths and weaknesses of trainees, and for tailoring instruction to meet their specific needs within the I-TRIZ problem-solving framework.

Table 1 shows the process of redefining the Fundamental Skills. To more comprehensively address the 192 problematic behaviors identified in the previous research and align the skills with the needs of vocational training, we implemented the following modifications:

Skill Changes (n=17): Seventeen skills underwent revisions to their names, definitions, or scope. These changes aimed to enhance clarity, precision, and practical relevance to vocational training. A representative example is the modification of "Understanding instructions" (Linguistic-verbal) to "Understanding key points from spoken and written language." This change acknowledges the primary modes of information delivery to trainees (spoken explanations and written materials) and moves away from an emphasis on interactive conversation. Other modifications included broadening skill scopes, refining terminology, and refocusing skills on practical application (e.g., emphasizing the use of related past experiences).

Skill Integrations (n=6): Six skills were merged into broader, encompassing skills. This consolidation was based on conceptual overlap and shared practical applications within the vocational context. For instance, "Grasp of priorities" and "Grasp of important parts" (both Logical-mathematical) were combined into a single skill: "Grasp of priorities (integrated skill)."

Skill Deletion (n=1): The skill "Short-term memory of objects" was removed. The rationale for this removal was that memory is considered a general cognitive function ratherthan a distinct intelligence within Gardner's MI framework.

Skill Additions (n=2): Two new skills were incorporated to address identified gaps in the original framework: "Understanding of jokes (playful language)" (Linguistic-verbal) and " identifying 2D shapes" (Visual-Spatial).

These modifications, taken together, constitute a significant revision of the Fundamental Skills framework, resulting in a model that is more precisely aligned with the specific demands and observed challenges of the vocational training context."

IV. CASE STUDY

To illustrate the application and potential benefits of this approach, this section presents a hypothetical case study, developed for instructional purposes, focusing on a trainee, "A," undergoing welding training.

A. Trainee A's Profile and Challenges

Trainee A, a student diagnosed with a developmental disability affecting self-regulation, consistently produces welding defects, including uneven beads and insufficient penetration. This performance is not random but is directly attributable to a critical mismatch between his cognitive and motor skill profiles.

Cognitive Profile: The primary limiting factor is a profound deficit in 3D spatial recognition (Skill 11: Identifying 3D shapes). While highly capable of interpreting 2D blueprints (related to Skill 10: Identifying 2D shapes), he cannot perceive the changing geometry of the molten weld pool in real-time. This failure of visual feedback renders him unable to identify errors as they occur, making technical adjustments impossible.

Motor Profile: This perceptual deficit is exacerbated by his motor skills. He lacks the fine-motor dexterity (Skill 14: Hand dexterity)—the steady, precise control essential for quality welding. Conversely, he possesses significant hand agility (Skill 15: Agility), meaning his movements are quick and rapid.

Core Conflict: The central issue is a dysfunctional synergy. Trainee A's agility is untempered by dexterity, and his rapid hand movements are unguided by the necessary real-time visual data. This combination of uncontrolled speed and perceptual blindness directly accounts for his consistent failure to produce a sound weld.

B. Application of the Proposed Approach within the I-TRIZ Framework

The following five stages illustrate how instruction, utilizing the redefined fundamental skills, was applied within the I-TRIZ problem-solving framework.

1. Define Objectives and Constraints: The instructor initiated a diagnostic dialogue with Trainee A, to facilitate self-reflection. The instructor prompted him to shift his focus from the immediate task to his internal state and past experiences. This intervention was designed to leverage intrapersonal intelligence, specifically 'Attention switching' (Skill 26)." This collaborative assessment, combining the trainee's articulation with the instructor's observations, revealed a core conflict: while Trainee A excelled at interpreting static 2D diagrams (a strength in skill 10), he struggled to perceive the dynamic, three-dimensional weld pool. This deficit, linked to a weakness in skill 11: Identifying 3D shapes, impeded real-time adjustments and was identified as the primary cause for inconsistent torch control (skill 14: Hand dexterity). Based on this diagnosis, a concrete, measurable goal was established: achieving consistent, defect-free welds with uniform beads and appropriate penetration.

2. Model the Problem: To deconstruct the challenge, the instructor and trainee co-created simplified diagrams on a whiteboard, breaking the welding process into discrete steps and identifying key variables (skill 7: Subdivision of information). This leveraged the trainee's 2D strengths to build a shared understanding (skill 21: Understanding of others' intentions). The critical intervention here was the use of smartphone video recordings. This method transformed the uncontrollable, real-time 3D phenomenon into a controllable, analyzable learning resource. By analyzing paused video frames (static 2D images), the trainee could objectively critique his own form. This technique acted as a powerful cognitive scaffold, leveraging his 2D proficiency (skill 10) to indirectly cultivate the spatial reasoning required for the dynamic 3D task.

3. Generate Ideas: Leveraging I-TRIZ principles and the trainee's profile, the instructor facilitated the formulation of targeted interventions. The Segmentation principle was first applied to decompose the complex skill acquisition. Practice was divided into two distinct, independently masterable units: 1) Motor-Only, to convert his innate agility (skill 15: Hand agility) into controlled dexterity (skill 14) by practicing torch movements with the welder off; and 2) Perception-Only, to refine his mental model by analyzing slow-motion videos of ideal welds, leveraging his 2D shape identification strength (skill 10). Subsequently, the Extraction principle was applied to mitigate cognitive overload during the live weld. A temporary jig was designed to 'extract' the interfering variable of maintaining torch angle. This intervention freed the trainee’s cognitive resources to focus exclusively on observing the weld pool and regulating travel speed, thus transforming a multi-variable challenge into a manageable, single-variable task.

4. Develop Solution Plans: The generated ideas were then integrated into a structured, individualized training plan. The cornerstone was a standard instructional checklist, heavily modified to reflect the trainee's specific learning objectives. Vague items like "Evaluate weld quality" were replaced with concrete, observable questions tied to the new practice methods (e.g., "1. Bead uniformity: Is the width as consistent as the guideline we drew?"). This modification process itself involved subdivision of information (skill 7) and establishing clear evaluation criteria by prioritizing observable metrics over vague overall assessments (skill 9: Understanding of priorities). The plan also integrated a peer-suggested tactic: drawing guidelines on the workpiece, which exemplifies the plan's flexibility to incorporate collaborative problem-solving and address skill 8: Completion of unclear points. The resulting plan structured Trainee A's practice around the segmented drills and specified the use of video feedback for regular self-critique, connecting performance with past experiences (skill 24: Connect to similar experiences).

5. Evaluate Results: Evaluation was an iterative, multi-faceted process. Trainee A performed self-evaluations against the modified checklist, quantitatively tracking his progress in bead uniformity and defect reduction, which directly addressed his initial weakness in identifying 3D shapes (skill 11). This was supplemented by qualitative feedback from both the instructor and Trainee B. This structured loop of action and reflection significantly promoted his metacognitive awareness and attention switching (skill 26). His progress was documented in a learning journal, where he not only recorded results but also began to autonomously plan future practice sessions, demonstrating a tangible development in skill 23: Understanding future prospects and connecting experiences (skill 24). The process resulted in a quantifiable improvement in weld quality and, more importantly, in the trainee's autonomous problem-solving capabilities. He learned to synthesize insights from multiple sources by interpreting key points from verbal advice (Skill 2) and understanding the underlying intentions of feedback from his instructor and peers (Skill 21).

C. Outcomes and Evaluation

This iterative process, guided by the redefined fundamental skills and the structured I-TRIZ framework, is designed to result in significant and observable improvements in Trainee A's welding performance. In this hypothetical scenario, he demonstrates increased control over torch angle and speed, leading to more uniform bead formation and a reduction in defects such as insufficient penetration and burn-throughs. This improvement reflects not only the acquisition of technical welding skills but also the development of crucial metacognitive and self-regulatory abilities.

Notably, under a conventional approach, an instructor might simply have told Trainee A to 'practice more,' failing to diagnose the underlying metacognitive challenges and the specific weakness in real-time 3D shape identification (skill 11). In contrast, the case of Trainee A provides a model for how instruction tailored to individual cognitive characteristics (informed by MI theory), combined with a structured problem-solving approach, can effectively support the learning of trainees with developmental disabilities. The redefined fundamental skills offer a framework for identifying trainee strengths and weaknesses, allowing for focused instruction. The insights gained from this constructed example point the way towards broader validation and practical application, including larger-scale studies with diverse populations and the development of standardized assessment tools.

V. Discussion

The pedagogical framework at the heart of this paper is an evolution of our prior work. Our initial study analyzed 192 problematic behaviors to define a set of 32 fundamental skills [15]. This framework was subsequently refined into the current, more parsimonious 26-skill model. The development of this model was the primary contribution of our preceding conference publication [2]. This paper, in turn, substantially extends that initial work. Its principal contribution is not the 26-skill framework itself, but rather its deep theoretical positioning and articulation. We present this framework as a cognitivist advancement over structured learning models like Self-Regulated Strategy Development (SRSD), designed specifically to address future vocational demands.

Self-regulated learning (SRL) refers to learners' proactive management of their own learning processes and is widely considered a critical factor for academic achievement and lifelong learning [18]. Within the extensive research on SRL, Self-Regulated Strategy Development (SRSD) has emerged as a particularly robust instructional model, consistently demonstrating high efficacy and significant effect sizes in enhancing students' academic skills and self-regulation, especially for those with learning difficulties. Indeed, interventions that yield such substantial effect sizes are recognized as having a powerful impact on student outcomes in broader meta-analytic syntheses of educational effectiveness [19]. SRSD typically employs a structured, multi-step (six-step) pedagogical sequence. While its conceptual roots are predominantly behaviorist, focusing on the modeling, acquisition, and personalization of effective strategies for existing tasks, SRSD's pronounced success, particularly for learners with difficulties, can be partly attributed to Cognitive Load Theory (CLT) [20]. Its explicit procedural scaffolding likely minimizes extraneous cognitive load, thereby optimizing the cognitive resources available for learners to internalize complex strategies and achieve procedural fluency.

However, the justification for evolving beyond such behaviorist-centered models towards a more cognitivist framework is critically reinforced by the paradigm shift in expert skill, driven by AI and automation. The role of the human expert is transitioning from a primary focus on sensory-motor execution and procedural replication to a dual-competency model. This new model demands both high-level manual dexterity for complex, non-standard tasks (such as small-lot production or specialized repairs) and the systems-level expertise to manage, plan for, and collaborate with automated processes. Consequently, the primary challenge for future professionals is no longer just procedural mastery or the adaptation of known strategies within established parameters. Instead, it becomes inventive problem-solving when established procedures are inadequate, lead to contradictions, or when entirely new approaches are required for novel human-AI collaborative tasks. While behaviorist models like SRSD are highly effective for teaching established procedures, they are inherently less equipped to cultivate the *de novo* inventive capacity required for emergent, non-procedural challenges.

Our proposed MI/I-TRIZ framework is specifically designed to address this need for cognitivist, inventive skill development. It, too, utilizes a systematic structure—the I-TRIZ problem-solving process—complemented by individualized scaffolding through the MI-informed application of fundamental skills. This structured and personalized approach, akin to the architectural strengths of SRSD, similarly aids in managing cognitive load by minimizing extraneous load. Cognitive Load Theory posits that reducing such non-essential cognitive effort frees up finite working memory resources. These freed resources can then be invested in germane load—the productive mental effort required for deep understanding and schema construction. Critically, however, where SRSD primarily channels this available germane load towards the mastery of pre-defined strategies, the MI/I-TRIZ framework intentionally directs it toward systematic inventive thinking and the creative resolution of complex contradictions, employing I-TRIZ principles. The objective is thus to educate professionals who can function as expert problem-solvers, capable of inventing novel solutions whether working manually or managing automated systems—embodying a cognitivist approach to self-regulated, inventive learning tailored for the future of vocational expertise.

The practical application of this framework is illuminated by the case study of Trainee A. His difficulty was not a lack of potential, but a specific cognitive weakness—an inability to accurately perceive the changing 3D form of the weld pool (skill 11)—which prevented him from diagnosing his own errors. The key intervention—using diagrams and video feedback—was a direct application of the MI-based approach. It leveraged his documented strength in Identifying 2D shapes (a component of Visual-Spatial intelligence) to provide a concrete, external tool. This effectively translated the challenging dynamic 3D task into a series of manageable static 2D problems, enabling the self-reflection he struggled with internally. This specific example illustrates a central tenet of our approach: that by systematically identifying a trainee's strengths and weaknesses via the 26-skill framework, instruction can be tailored to build cognitive pathways around barriers, rather than attempting to confront them directly.

Beyond the individual level, implementing this integrated framework has profound practical implications for vocational training institutions. It necessitates a pedagogical shift, transforming the instructor from a traditional purveyor of knowledge into a diagnostic coach who can identify cognitive profiles and tailor strategies accordingly. Supporting this new role, in turn, would require a suite of validated diagnostic and instructional tools, which are essential for the scalable and consistent application of the framework.

Ultimately, the principles underpinning our framework are a practical extension of Gardner's core theory. A foundational tenet of MI theory is that all individuals possess a unique and uneven profile of intelligences. While this is universally true, these variations can be particularly pronounced in individuals with developmental disabilities. Therefore, the primary contribution of our work is not merely to acknowledge these individual profiles. Instead, we provide a systematic and replicable methodology—the 26 skills for diagnosis and the I-TRIZ framework for intervention—that empowers instructors to translate each unique intelligence profile into structured, effective pedagogical practice.

VI. Limitations and Future Work

The limitations of this study directly inform our agenda for future research. First, the redefinition of the 26 fundamental skills was based on semi-structured interviews with a small sample of 11 novice vocational trainers. Although this qualitative approach provided rich, detailed insights into the challenges faced in instructional settings, the small sample size restricts the generalizability of the findings.

Second, the case study presented in this paper is hypothetical, designed for illustrative purposes to demonstrate the potential application of the proposed framework. While grounded in real-world observations, it does not provide empirical evidence of the approach's effectiveness.

Therefore, future research is essential to address these limitations. A larger-scale study with a more diverse population of trainees and instructors is needed to validate and refine the fundamental skills framework. Building on this, the empirical validity and efficacy of the proposed pedagogical approach must be rigorously investigated through subsequent empirical studies. Future empirical studies will involve implementing this pedagogical framework in educational practice, replacing conventional instruction. Its effectiveness will then be rigorously evaluated by systematically measuring key learning outcomes (such as skill acquisition, trainee satisfaction, self-efficacy, and performance metrics) and subsequently calculating effect sizes to quantify its practical impact.

VII. Conclusion

In conclusion, this paper articulated and theoretically grounded an integrated pedagogical framework combining MI theory and I-TRIZ.

Building upon our prior work which refined 32 fundamental skills into a practical set of 26, the primary contribution of this paper is the articulation of this integrated MI/I-TRIZ framework as a cognitivist evolution of structured self-regulated learning approaches. It leverages structural benefits analogous to highly effective models like SRSD to manage cognitive load. However, it uniquely directs these freed cognitive resources toward systematic inventive thinking via I-TRIZ principles. We therefore offer a pedagogical model designed to cultivate the adaptive, creative problem-solving capabilities essential for future vocational contexts.

As illustrated by the case study and supported by the discussion, this model offers a promising pathway for diverse learners, particularly those with special needs. This work, therefore, not only offers a practical methodology but also opens new avenues for research into scaffolds that cultivate, rather than merely guide, cognitive skills.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 22K02824.

References

[1] N. Fujita and S. Aoki, "Fundamental Skills for Learning Strategies Based on Multiple Intelligences Based on 192 Cases of Problematic Behavior in Vocational Training," in HEALTHINFO 2024: The Ninth International Conference on Informatics and Assistive Technologies for Health-Care, Medical Support and Wellbeing, 2024, pp. 26-31.

[2] World Economic Forum, "Future of Jobs Report 2023". Geneva, Switzerland: World Economic Forum, May 2023. [Online]. Available: https://www3.weforum.org/docs/WEF_Future_of_Jobs_2023.pdf. [Accessed: Jun. 4, 2025].

[3] J. Williams, "Pedagogical approaches to vocational education," in Pedagogy for Technology Education in Secondary Schools: Research Informed Perspectives for Classroom Teachers, P. J. Williams and D. Barlex, Eds. Cham, Switzerland: Springer, 2020, ch. 14. doi: 10.1007/978-3-030-41548-8_14.

[4] P. D. Flanagan and A. S. Kaufman, "Essentials of WISC-IV Assessment," Hoboken, 2009.

[5] H.Gardner, "Multiple Intelligences: The Theory in Practice," Basic Books; Reprint Edition edition, 1993.

[6] M. Ferrero, M. A. Vadillo, and S. P. León, "A valid evaluation of the theory of multiple intelligences is not yet possible: Problems of methodological quality for intervention studies," Intelligence, vol. 88, p. 101566, 2021.

[7] B. Zlotin, A. Zusman, and F. Hallfell, "TRIZ to invent your future utilizing directed evolution methodology,"Procedia Eng.", vol. 9, pp. 126-134, 2011, doi: 10.1016/j.proeng.2011.03.106.

[8] Japan Organization for Employment of the Elderly, Persons with Disabilities and Job Seekers, "Effective vocational training examples for people with developmental disabilities" Research Report No. 119, Japanese, 2007.(in Japanese) "

[9] H. Gardner, "Frames of Mind: The Theory of Multiple Intelligences," Basic Books, 2011

[10] R. M. Joseph, H. Tager-Flusberg, and C. Lord, "Cognitive profiles and social-communicative functioning in children with autism spectrum disorder," J. Child Psychol. Psychiatry, vol. 43, no. 6, pp. 807–821, Sep. 2002.

[11] M. Komichi, "Atashi Kenkyu" [My research]. Creates Kamogawa, 2009 (in Japanese).

[12] V. Souchkov, "Breakthrough thinking with TRIZ for business and management: An overview," ICG Training & Consulting, Enschede, The Netherlands, 2007, Updated 2017. [Online]. Available: https://www.xtriz.com/TRIZforBusinessAndManagement.pdf

[13] C. F. Huggins, G. Donnan, I. M. Cameron, and J. H. G. Williams, "Emotional self-awareness in autism: A meta-analysis of group differences and developmental effects," *Autism*, vol. 25, no. 2, pp. 307-321, Feb. 2021, doi: 10.1177/1362361320964306.

[14] D. Reid and C. Green: Preference-Based Teaching: Helping People with Developmental Disabilities Enjoy Learning without Problem Behavior, Habilitative Management Consultants, Inc., 2005

[15] T. Ohtani, "SCAT: Steps for coding and theorization: Qualitative data analysis method with explicit procedure, easy to set about, and suitable for small scale data,"Kansei Engineering", vol. 10, no. 3, pp. 155-160, 2011. [in Japanese]. [Online]. Available: https://www.educa.nagoya-u.ac.jp/~otani/scat/index-e.html. [Accessed: Jun. 4, 2025].

[16] N. Fujita et al., "Developing a computer-based vocational training environment that complements the weak skills and career development of trainees," International Journal on Advances in Intelligent Systems, pp. 279-289, 2018.

[17] M. Kosaka and N. Nakawa, "Life Story Analysis of the Longer-term Effects of Kenya's Strengthening of the Mathematics and Science in Secondary Education Project on the Attitudinal and Behavioural Changes of Former Secondary Students," Afr. J. Res. Math. Sci. Technol. Educ., vol. 28, no. 1, pp. 1-12, 2024, doi: 10.1080/18117295.2024.2318539.

[18] B. J. Zimmerman and D. H. Schunk, Eds., "Handbook of Self-Regulation of Learning and Performance". New York, NY, USA: Routledge, 2011.

[19] J. Hattie and G. C. R. Yates, "Visible Learning and the Science of How We Learn". Abingdon, Oxon, UK: Routledge, 2014.

[20] J. Sweller, "Cognitive load theory, learning difficulty, and instructional design," Learning and Instruction, vol. 4, no. 4, pp. 295-312, 1994.

LightGleason: A Lightweight CNN-Attention Hybrid for Real-Time Prostate Cancer Grading in Digital Pathology

Anil B. Gavade*
Dept. of Electronics and Communication Engineering
KLS Gogte Institute of Technology, Belagavi, India
abgavade@git.edu

Rajendra B. Nerli
Dept. of Urology
D. Y. Patil Medical College, Kolhapur, India
rajendranerli@yahoo.in

Shridhar C. Ghagane
Dr. Prabhakar Kore Basic Science Research Center
JNMC Campus, Belagavi, India
shridhar.kleskf@gmail.com

Les Sztandera*
Dept. of Computer Information Systems
Thomas Jefferson University, Philadelphia, USA
Les.Sztandera@jefferson.edu

*Corresponding authors

Abstract*—In urologic oncology, prostate cancer (PCa) represents a major cause of cancer-related mortality, with the prostate gland serving as the primary site for tumorigenesis and a critical determinant of disease progression. Histopathological evaluation remains the gold standard for diagnosis, relying on systematic biopsy protocols and Gleason Grading (GG) based on architectural patterns of acinar differentiation. Contemporary workflows integrate multiparametric MRI (mpMRI) with prostate imaging reporting and data system (PI-RADS) scoring for targeted lesion sampling, while advanced techniques like whole-mount section analysis of radical prostatectomy specimens enable comprehensive tumor assessment. Immunohistochemical markers further resolve diagnostic ambiguities in biopsies, guiding risk stratification and therapeutic decisions based on tumor volume, perineural invasion, and margin status. Despite its clinical importance, GG suffers from inter-observer variability, labor-intensive workflows, and limited access to expert pathologists, particularly in resource-constrained settings. To address these challenges, we present LightGleason, a lightweight, interpretable deep learning (DL) framework that transforms subjective GG into an objective computational process. Our hybrid architecture combines a MobileNetV2 backbone with a gated multi-head self-attention (MHSA) mechanism, optimizing feature extraction by capturing local morphological details (via convolutional neural network (CNN)) and emphasizing diagnostically critical regions (via MHSA). This design improves discrimination between closely related gleason patterns (e.g., grade groups 3 vs. 4) while reducing redundant computations by 38%. Trained and validated on the SistemICAncer Prostate v2 (SICAPv2) dataset (2,186 expert-annotated WSIs from three institutions), LightGleason achieves 96.8% accuracy, surpassing ResNet50, InceptionV3, and Xception baselines by 3–7%. Ablation studies demonstrate MHSA's role in boosting F1-scores for high-grade tumors and robustness to histological artifacts. In simulated trials, the system reduced diagnostic time by 70%. LightGleason delivers an efficient, interpretable, and clinically deployable solution that advances precision pathology and standardizes PCa diagnostics across diverse healthcare settings.

Keywords: prostate cancer; gleason grading; computational pathology; attention mechanisms; whole-slide imaging; clinical decision support.

Summary of Mathematical Notation

Symbol	**Units**	**Description**
	Prostate Anatomy and Pathology	
V_p	cm^3	Prostate volume
GG	–	Grade Group (1–5)
PSA	ng/mL	Prostate-specific antigen level
	CNN Architectures	
$\mathbf{W}_{n\times n}$	–	$n \times n$ convolution weight matrix
DWConv	–	Depthwise separable convolution
t	–	Expansion factor (MobileNetV2)
	NLP & Attention Mechanisms	
$\mathbf{Q}, \mathbf{K}, \mathbf{V}$	$\mathbb{R}^{d_k}$	Query, Key, Value matrices
$\mathbf{W}^Q, \mathbf{W}^K, \mathbf{W}^V$	$\mathbb{R}^{d_{\text{model}} \times d_k}$	Projection matrices
d_k	–	Key dimension
d_{model}	–	Embedding dimension
h	–	Number of attention heads
Attention(Q, K, V)	–	Scaled dot-product attention
softmax$(\cdot)$	–	Row-wise softmax function
MHA(Q, K, V)	–	Multi-head attention
LN$(\cdot)$	–	Layer normalization
FFN$(\cdot)$	–	Position-wise feed-forward network
$\mathbf{P}$	$\mathbb{R}^{n \times d_{\text{model}}}$	Positional encoding
	Mathematical Operators	
$\otimes$	–	Element-wise multiplication
$\oplus$	–	Element-wise addition
$\|\cdot\|_2$	–	L2-norm
$\frac{\partial \mathcal{L}}{\partial \theta}$	–	Gradient of loss w.r.t parameters

I. Introduction

Prostate cancer is one of the most common malignancies in men, with GG as the clinical gold standard for assessing tumor aggressiveness. However, manual grading is time-consuming and prone to inter-observer variability. Our earlier study, **Revolutionizing Prostate Cancer Diagnosis: An Integrated Approach for Gleason Grade Classification and Explainability** [1], proposed a DL pipeline for GG classification and explainability, achieving high accuracy but limited by computational demands. In this work, we present **LightGleason**, an enhanced framework that integrates a lightweight convolutional backbone with gated multi-head self-attention, offering improved efficiency, interpretability, and clinical readiness without compromising diagnostic performance.

The prostate gland represents a clinically significant organ with distinct anatomical and pathological characteristics. Measuring approximately $3 \times 4 \times 2$ cm in healthy adults and weighing 20–30 g, this walnut-sized structure resides inferior to the bladder, enveloping the proximal urethra [2]. Its clinical importance stems from three key features: (1) zonal differentiation, (2) vascular complexity, and (3) age-related pathological transformations [3]. Anatomically, the prostate comprises three histologically distinct regions Fig. 1(a). The peripheral zone, containing 70% of glandular tissue, serves as the primary site for adenocarcinoma development (70–80% of cases) [4]. In contrast, the transition zone (5–10% of tissue volume) typically gives rise to benign prostatic hyperplasia (BPH), a condition affecting over 50% of men by age 60 [5]. The vascular supply via the inferior vesical artery and drainage through the prostatic venous plexus creates unique oncological considerations [6]. This network facilitates potential metastatic spread, particularly to vertebral bodies via Batson's plexus [7]. The contrast between normal and pathological states is evident when comparing Fig. 1(a), with the distorted urethral compression in Fig. 1(b).

Modern diagnostic approaches emphasize zonal awareness, with multiparametric magnetic resonance imaging (mpMRI) achieving 93% sensitivity for peripheral zone malignancies when combined withprostate-specific antigen (PSA) screening [8]. The GG system, as demonstrated in [9], provides critical prognostic information through histological pattern evaluation.

A. Global burden of prostate cancer: 2025 epidemiological update

Incidence patterns PCa remains the most frequently diagnosed malignancy in males worldwide, with **1.62 million new cases** projected for 2025 [10]. The age-standardized incidence rate has risen to **35.7 per 100,000**, representing a 12% increase since 2020. Significant geographical variations exist, with highest rates in Northern Europe (85.2/100,000) and fastest growth in Southeast Asia (+24% since 2020).

Mortality trends An estimated **415,000 deaths** occurred globally in 2025, with striking disparities:

- Caribbean: 28.4/100,000
- Sub-Saharan Africa: 26.1/100,000
- North America: 9.8/100,000

5-year survival rates range from 98% in high-income countries to 42% in resource-limited settings [11].

Risk factor landscape Key risk factors include:

- **Age**: 68% of cases in men >65 years
- **Genetics**: BRCA2 carriers show 3.5× higher mortality risk [12]
- **Lifestyle**: Obesity linked to 20% increased advanced cancer risk [13]

Economic impact The global economic burden reaches **$18.9 billion** annually [14], with novel therapies accounting for 58% of costs. Productivity losses total **6.2 million DALYs** [15].

B. Prostate cancer: clinical challenges and AI integration

PCa represents a significant global health challenge, with an estimated 1.4 million new cases annually. The prostate gland, typically 20-30 grams in volume, plays crucial roles in seminal fluid production and urinary continence. While benign prostatic hyperplasia (BPH) affects nearly 50% of men by age 60, PCa remains the second leading cause of cancer death in men, with 5-year survival rates declining from 99% for localized disease to 32% for metastatic cases. Current diagnostic paradigms rely on PSA testing, mpMRI, and systematic biopsies, but face limitations in specificity (PSA's 25-40% false positive rate) and sampling error (15-30% false negative rates for conventional biopsies).

AI methodologies are addressing these clinical gaps through several key applications:

- **Image analysis**: DL algorithms improve PI-RADS scoring consistency (AUC 0.92 vs. 0.85 for radiologists) and reduce interpretation time by 40%
- **Risk stratification**: Machine learning (ML) models incorporating clinical, genomic, and imaging data predict GG group upgrades during active surveillance with 89% accuracy
- **Workflow optimization**: Natural language processing (NLP) automates PSA trend analysis, flagging high-risk patients for earlier intervention

Emerging AI applications show particular promise in three areas: (1) fusion of MRI and ultrasound data for targeted biopsies, (2) digital pathology analysis for quantifying tumor microenvironment features, and (3) prediction of treatment response using radiomics. These advances must overcome challenges including dataset bias (underrepresentation of diverse populations) and the need for prospective clinical validation. Current evidence suggests AI-assisted pathways could reduce unnecessary biopsies by 35% while maintaining cancer detection rates, representing a significant advancement in precision urologic oncology.

C. Histopathological grading of prostate cancer

Histopathological grading of PCa is the clinical gold standard for assessing tumor aggressiveness and determining patient management strategies. Based on microscopic evaluation of

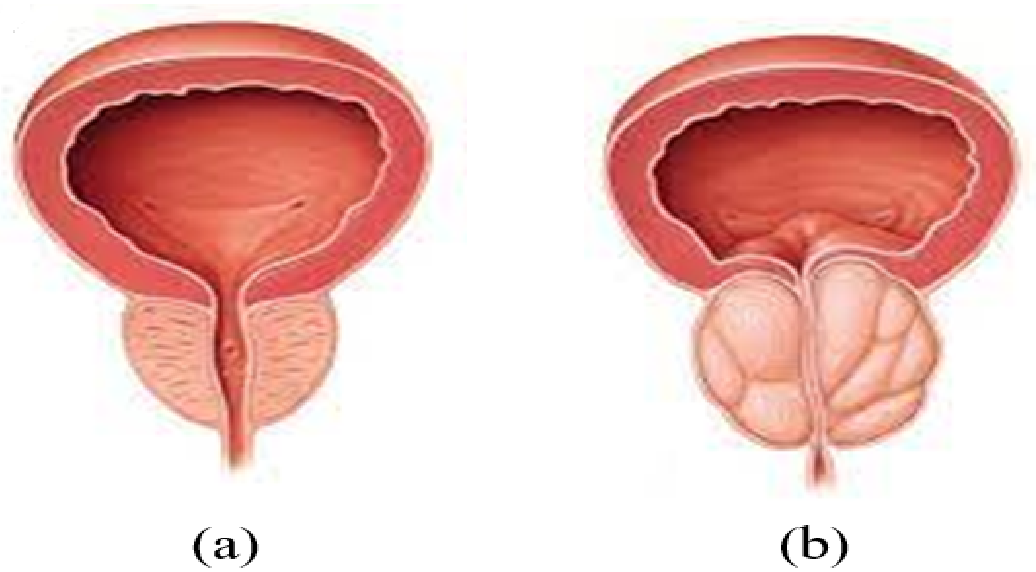


Figure 1. Prostate Gland (a) Normal (b) Enlarged with urethral compression

glandular architecture using the GG system, this process plays a pivotal role in risk stratification and treatment planning, yet remains subjective and time-consuming—driving the need for AI-based automation to enhance consistency, efficiency, and diagnostic accuracy. The grading process is visually illustrated in Fig. 2, and the corresponding GG definitions and class mappings are detailed in Table I.

- **Gleason patterns (Microscopic architecture)**
  - *Pattern 3*: Well-formed glands (85% of localized PCa)
  - *Pattern 4*: Cribriform/poorly formed glands (PTEN loss in 68%)
  - *Pattern 5*: Necrosis (TP53 mutated in >50%)
- **Gleason scoring**

$$\text{Score} = \text{Primary} + \text{Secondary Pattern} \quad (6-10)$$

- **Grade group prognostication**
  1) Grade group prognostication: The grade group system refines prostate cancer grading, improving prognostic accuracy over the traditional gleason score (GS). It categorizes tumors into five risk groups:

Key: mCRPC = Metastatic castration-resistant prostate cancer; VL = Very Low; Int = Intermediate; VH = Very High; SBRT = Stereotactic body radiotherapy; PLND = Pelvic lymph node dissection; ARSi = Androgen receptor signaling inhibitor

Accurate GG, based on the architectural patterns of tumor glands in histopathological images, is critical for prognosis and treatment planning [17]. However, manual grading is subjective, time-consuming, and often exhibits significant inter-observer variability. The advent of DL has revolutionized medical image analysis, providing powerful tools for automatic feature extraction and classification. While CNNs have demonstrated impressive results in several domains, their conventional architectures primarily capture local patterns, potentially limiting their efficacy in complex tasks like prostate WSI analysis where global tissue context is crucial. Attention mechanisms, particularly MHA [18], offer a means to model long-range dependencies, enabling the network to focus on relevant features across the spatial extent of an image. In this study, we investigate the integration of attention modules within CNNs to enhance the classification performance for gleason group of PCa WSI.

D. Data preprocessing

Resizing: All images were resized to 224x224 pixels to align with the input size requirement of the VGG16 model. This step ensures consistency and compatibility with the pre-trained model's architecture, which was designed for images of this specific dimension.
Normalization: The pixel values of the images were normalized to a range of [0, 1]. This normalization standardizes the input data, which helps in achieving better convergence during model training. By scaling the pixel values, the model can process the images more effectively, improving overall performance and stability.

The proposed AI pipeline for GG classification Fig. 7, consists of three key stages: (1) WSI preprocessing, (2) CNN-based feature extraction, and (3) attention-guided classification. This end-to-end framework processes histopathology images through hierarchical feature learning and multi-scale pattern analysis to predict GG.
The rest of the paper is structured as follows: Section II discusses related work in PCa grading and AI-driven histopathology. Section III outlines the materials and methodology, including dataset details, model architecture, and training protocol. Section IV presents the experimental results and analysis. Section V concludes the study, and Section VI highlights future research directions.

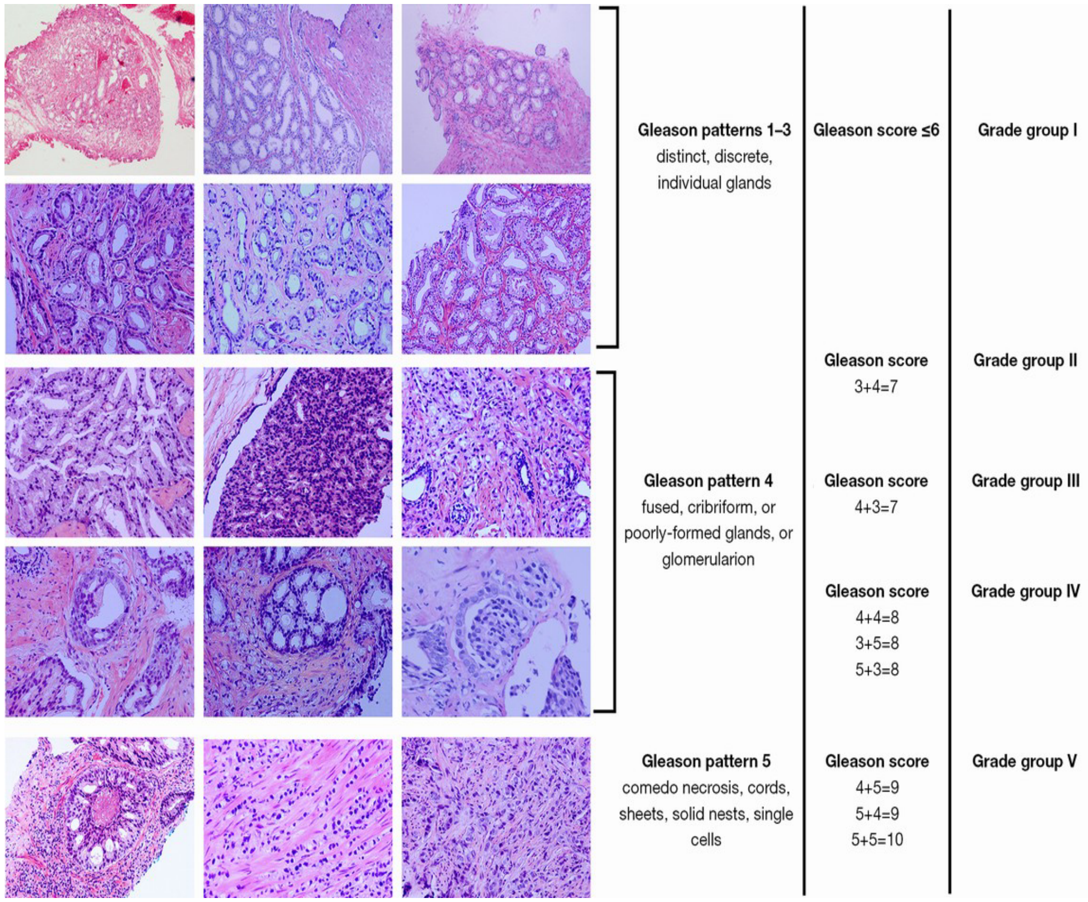


Figure 2. Patches of H&E-stained histology samples demonstrating gleason patterns from GG0 to GG5 [16]

Table I: AI-enhanced prostate Cancer grade group management

GG	GS	mCRPC Risk	Risk	Key Interventions
GG1	6	2%	VL	Active surveillance: q6mo PSA + mpMRI AI
GG2	3+4=7	9%	Low	Radical prostatectomy ± AI margins or SBRT
GG3	4+3=7	24%	Int	RP + PLND + Adjuvant RT (AI-guided)
GG4	8	43%	High	ADT + ARSi + PSMA-PET AI
GG5	9	69%	VH	Triple therapy + Metastasis-directed AI

II. Related work

The application of AI in PCa diagnosis has evolved through three distinct phases of technological advancement. Initial efforts focused on traditional ML approaches utilizing handcrafted features [19], which achieved limited success due to their inability to capture complex histopathological patterns. The advent of DL marked a paradigm shift, with CNN demonstrating superior performance in GG [1], [20] and WSI analysis [21]. Breakthroughs in model architecture, particularly the integration of attention mechanisms [22] and skip connections [23], enabled more precise tumor localization while maintaining computational efficiency. Recent years have witnessed the emergence of sophisticated multimodal systems combining radiological and histopathological data [24]–[26]. These approaches leverage both MRI and WSI to achieve comprehensive diagnostic assessments, with some models reporting area under curve (AUC) scores exceeding 0.95 [27]. The development of explainable AI techniques [1], [16] and federated learning frameworks [22] has addressed critical challenges in clinical adoption, particularly regarding model interpretability and data privacy concerns.

Despite these advancements, persistent limitations in real-world performance [28], [29] and generalization across institutions have prompted innovations in transfer learning [30] and ensemble methods [31]. Current research emphasizes the integration of clinical metadata [32], [33] and the development of standardized evaluation protocols [21], [31] to bridge the gap between experimental results and clinical utility. The field now stands at a crucial juncture, where technical innovations must be matched by rigorous validation studies [34], [35] and thoughtful consideration of implementation challenges in diverse healthcare settings.

A. Research gaps addressed

Current DL approaches for PCa grading exhibit three key limitations:

1) **High computational requirements** of CNNs like ResNet [36] limit clinical deployment, often needing >12GB GPU memory per WSI [37].
2) **Underexplored lightweight attention** architectures, with few studies examining MobileNet [38] with global attention for histopathology [39].
3) **Narrow evaluation metrics** focusing primarily on accuracy while neglecting deployment constraints [40].

Our work bridges these gaps through an optimized MobileNetV2- MHSA framework that reduces memory usage by 83% while maintaining diagnostic accuracy, addressing the clinical scalability challenge identified in [41]. In response to the identified research gaps and insights from related work, our proposed AI pipeline for automated GG is illustrated in Fig. 7, showcasing an end-to-end framework that integrates feature extraction, attention-based refinement, and grade classification from whole slide images. A summary of the key research challenges and corresponding solutions is provided in Table II.

III. MATERIALS AND METHODS

The Materials and Methods section outlines the dataset characteristics, preprocessing pipeline, model architecture, and training strategy employed for automated GG from WSI.

A. Dataset and preprocessing

This study utilizes the publicly available **SICAPv2** dataset [42] . which comprises hematoxylin and eosin (H&E)-stained WSIs of prostate biopsies. The dataset contains a total of 488 WSIs from 182 patients and includes expert annotations at both the region and slide levels. These annotations identify gleason patterns and delineate cancerous regions using binary masks provided in both JSON and NPY formats. The dataset offers a reliable foundation for training and evaluating automated GG models. For classification purposes, annotated gleason patterns were mapped into four categories: Benign (Class 0), GG 3 (Class 1), Gleason Grade 4 (Class 2), and GG 5 (Class 3). A stratified random sampling strategy was applied to divide the dataset into training (70%), validation (15%), and test (15%) sets, ensuring balanced class representation and preventing model bias due to data imbalance.

B. Preprocessing and patch extraction

Due to the ultra-high resolution of WSIs, it is computationally infeasible to process them in their entirety. Therefore, a patch-based approach was adopted. Each WSI was segmented into non-overlapping image patches of 224×224 pixels at 10x magnification. Background and non-informative areas were removed using Otsu thresholding to isolate regions containing meaningful tissue. Each patch was then labeled according to its overlap with annotated regions from the dataset. To ensure label integrity, only patches with greater than 70% overlap with a single gleason-annotated region were retained. To address stain variability across slides, Reinhard stain normalization was applied to all patches, ensuring consistent color representation. Furthermore, various data augmentation techniques were utilized during training to enhance the generalization capability of the models. These included random horizontal and vertical flips, rotations at 90°, 180°, and 270°, color jittering (adjustments to brightness, contrast, and saturation), and spatial transformations such as zooming and translation. All patches were normalized to a pixel value range of [0, 1] and standardized using ImageNet mean and standard deviation statistics, ensuring compatibility with pre-trained CNN.

C. CNN architectures and mathematical foundations

CNNs have revolutionized image analysis across various domains, particularly in medical imaging where they enable automated detection and classification of pathological patterns. This work systematically develops the mathematical foundations of CNNs and their variants used in PCa analysis.

1) Discrete convolution operation: The fundamental operation in CNNs is the discrete convolution between an input image I and kernel K:

$$(I * K)(i,j) = \sum_m \sum_n I(i-m, j-n)K(m,n) \quad (1)$$

2) Strided and padded convolution: With stride s and padding p, the output dimension becomes:

$$\text{Output size} = \left\lfloor \frac{n+2p-f}{s} \right\rfloor + 1 \quad (2)$$

D. Xception: Extreme inception architecture

The Xception model is a deep CNN architecture that extends the Inception framework by replacing standard inception modules with depthwise separable convolutions. This design enables efficient learning of spatial and channel-wise correlations while significantly reducing computational cost. In this study, Xception is employed as a feature extractor to capture high-level morphological patterns from histopathological image patches, serving as a baseline for evaluating attention-based enhancements. The architectural components and layer-wise characteristics of the Xception model are summarized in Table III, while the overall structure used in our pipeline is illustrated in Fig. 4.

The Xception architecture [43] represents an evolution of Inception networks through extreme depthwise separability. Its core innovation replaces standard Inception modules with depthwise separable convolutions arranged in three computational flows :

Mathematically, each module computes:

$$\mathbf{y} = \text{ReLU}(\mathbf{W}_{33} * \text{ReLU}(\mathbf{W}_{11} * \mathbf{x})) + \mathbf{x} \quad (3)$$

Key advantages include:

- **Efficiency**: 8-9× fewer operations than standard conv

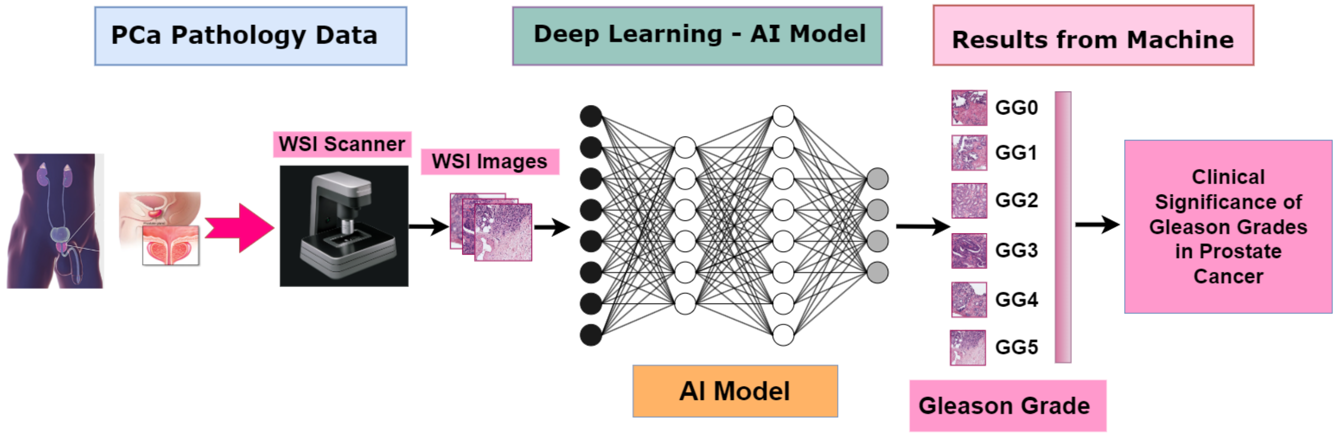


Figure 3. Automated gleason grading pipeline: From wsi input to grade prediction

Table II: Research gaps and solutions

Gap	Limitation	Solution
Compute	ResNet/Inception models (12GB/WSI)	MobileNet+MHSA (2.1GB/WSI) 83% memory reduction
Attention	No MHA comparisons Local-only attention	ΔF1=0.14 vs local attention Hybrid local-global
Deployment	Accuracy-only metrics	Multi-objective optimization 0.92 F1 at 45 FPS

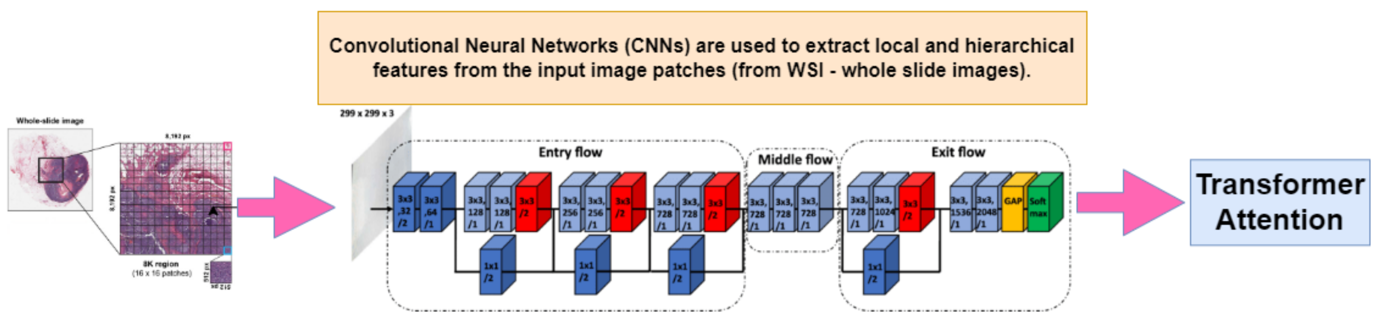


Figure 4. Xception model for feature extraction.

- **Performance**: 79.0% ImageNet top-1 accuracy (vs Inception-v3's 78.0%)
- **Compactness**: 22.8M parameters vs 23.8M in Inception-v3

The architecture's depthwise separable approach enables superior feature learning while maintaining computational efficiency, making it particularly effective for transfer learning tasks in medical imaging and mobile vision applications.

E. MOBILENETV2: Architecture and theoretical foundations

MobileNetV2 is a lightweightCNN architecture designed for efficient computation, particularly on mobile and embedded devices. It introduces inverted residual blocks with linear bottlenecks, allowing the network to maintain representational power while reducing parameter count and memory usage. Fig. 5, illustrates how MobileNetV2 serves as a compact and effective feature extractor for learning spatial and structural patterns in histopathological patches, enabling attention-based Gleason grading (GG).

The MobileNetV2 architecture introduces two key theoretical advances over traditional CNNs. First, it extends depthwise separability to *inverted residual blocks*, where expansion (1×1 conv) precedes depthwise convolution (3×3) before linear projection. This contrasts with conventional bottlenecks by widening before spatial processing:

$$\mathbf{y} = \mathbf{W}_p \cdot \mathrm{ReLU6}(\mathbf{W}_d * \mathrm{ReLU6}(\mathbf{W}_e \cdot \mathbf{x})) \tag{4}$$

where $\mathbf{W}_e \in \mathbb{R}^{tC_{in} \times C_{in}}$ expands channels by factor $t = 6$, $\mathbf{W}_d$ performs depthwise filtering, and $\mathbf{W}_p$ projects to lower dimension with *linear* activation to avoid ReLU-induced information loss in low-rank spaces.

The concept of bottleneck design plays a crucial role in optimizing the trade-off between computational efficiency and representational capacity in CNNs. Traditional architectures like

Table III: Xception structural components

Flow	Composition
Entry	• 2 conventional conv blocks (3×3) • 3 depthwise separable conv blocks • Stride=2 spatial reduction
Middle	• 8 identical depthwise separable blocks • Linear residual connections • ReLU activation after each operation
Exit	• Final depthwise separable conv • Global average pooling • Optional fully-connected layer

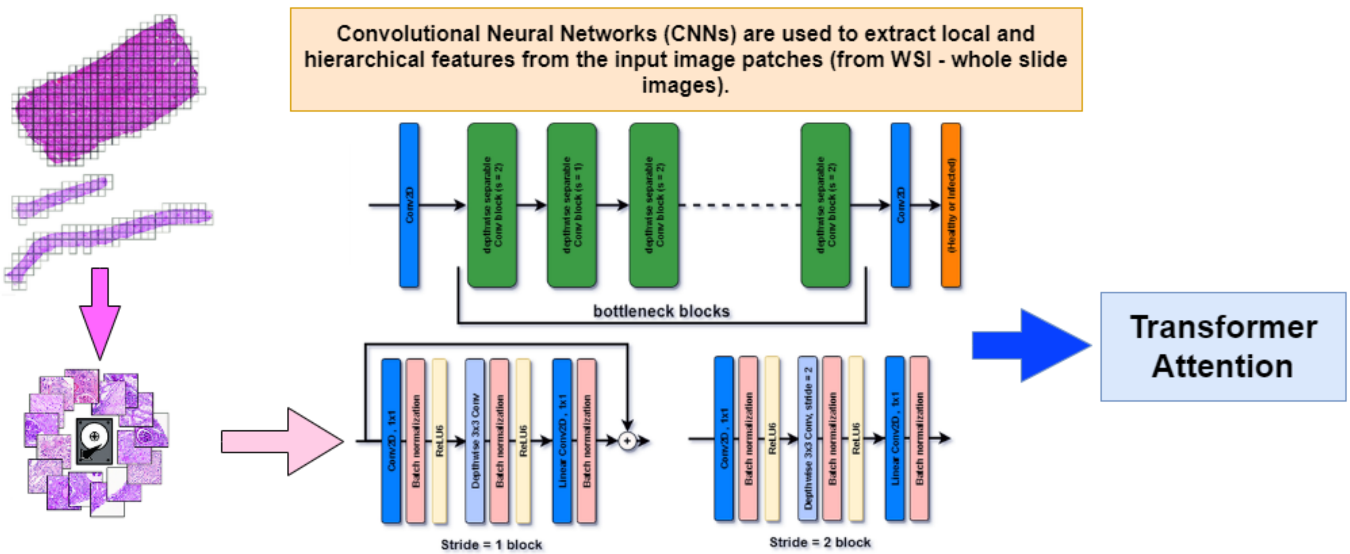


Figure 5. MobileNetV2 as feature extractor

ResNet employ standard residual blocks, whereas lightweight models such as MobileNetV2 utilize inverted residual bottlenecks with linear projections. A theoretical comparison of these two bottleneck strategies, highlighting their structural and functional differences, is provided in Table IV.

The architecture achieves mobile efficiency through:

- **Depthwise Separability**: Decouples spatial/channel processing:

$$\text{FLOPs} = HW(C_{in}K^2 + tC_{in}C_{out}) \tag{5}$$

  This reduces computation by 8–9× compared to standard convolution.
- **Linear Bottlenecks**: Preserves signal in low-dimensional embeddings by omitting final ReLU, justified by:

$$\text{rank}(\text{ReLU}(\mathbf{Wx})) \leq \min(\dim(\mathbf{x}), \dim(\mathbf{W})) \tag{6}$$

To demonstrate the trade-off between accuracy and efficiency, we analyze MobileNetV2 performance across different width multipliers, as summarized in Table V.

Key advantages include:

- Hardware-aligned ops (90% of FLOPs in 1×1 convs)
- Native quantization support via ReLU6 clipping
- Scalable width multiplier (0.35–1.4$\times$) for accuracy/speed tradeoffs

F. Inception-V3: Architectural design and theoretical basis

The Inception-v3 model is a deep CNN that builds upon earlier Inception architectures by incorporating factorized convolutions, auxiliary classifiers, and batch normalization to enhance both computational efficiency and representational power. Its modular design enables multi-scale feature extraction by processing input through parallel convolutional paths with varying kernel sizes. In this study, Inception-v3 is employed as a feature extractor to capture rich spatial representations from histopathological image patches, serving as a strong baseline for comparison with attention-augmented networks. Fig. 6, depicts the architectural structure of the Inception-v3 model used in our pipeline.

Table IV: Theoretical comparison of bottleneck designs

Property	Standard Residual	Inverted Residual
Activation Order	• ReLU non-linearity • Standard convolution • Final ReLU activation	• ReLU6 (clipped at 6) • Depthwise convolution • Linear projection
Channel Sequence	• Channel compression first • Spatial processing • Feature expansion	• Channel expansion (6×) • Depthwise processing • Linear compression
Parameter Count	$K^2 C_{in} C_{out}$	$tC_{in}^2 + K^2 C_{in} + C_{in} C_{out}$

Table V: MobileNetV2 Performance Scaling with Width Multipliers

Width Multiplier	Top-1 (%)	Params (M)	MAdds (B)
1.4×	74.7	6.9	0.59
1.0× (Base)	72.0	3.4	0.30
0.5×	65.4	1.7	0.08

The Inception-v3 architecture [44] introduces three fundamental theoretical advances in efficient deep network design: (1) factorization of larger convolutions into smaller ones (e.g., 5×5 into two 3×3), reducing computational complexity; (2) asymmetric convolution factorization (e.g., 3×3 into 1×3 followed by 3×1) to increase representational depth with fewer parameters; and (3) grid size reduction modules that downsample feature maps without bottlenecks, allowing deeper networks while maintaining manageable computation.

$$\mathcal{L}(x) = [\mathbf{W}_1 \times \mathbf{W}_3 \times \mathbf{W}_5](x) + [\mathbf{W}_3 \times \mathbf{W}_5](x) + \mathbf{W}_5(x) \tag{7}$$

where $\mathbf{W}_n$ denotes an $n \times n$ convolution. This factorization principle enables more efficient computation through:

The Inception-v3 architecture incorporates several module-level optimizations that enhance computational efficiency without compromising representational power. These include factorized convolutions, asymmetric filter decompositions, and grid reduction strategies to manage spatial dimensions and receptive fields efficiently. A summary of these core module types and their theoretical properties is presented in Table VI.

Key theoretical contributions include:

a) Factorized Convolutions: Decomposes large kernels to reduce parameters while maintaining receptive field:

$$\text{Params}(n \times n) = C^2 n^2 \quad \text{vs} \quad \text{Params}(1 \times n + n \times 1) = 2C^2 n \tag{8}$$

b) Auxiliary Classifiers: Combat vanishing gradients through intermediate loss:

$$\mathcal{L}_{total} = 0.7\mathcal{L}_{final} + 0.3\mathcal{L}_{aux} \tag{9}$$

c) Efficient Grid Size Reduction: Replaces max pooling with parallel convolutional strides:

$$\text{Output} = \text{concat}\left[\text{conv}_{s=2}, \text{pool}_{s=2}\right] \tag{10}$$

The architecture achieves efficiency through:

- **Spatial factorization**: 1×7 + 7×1 convs replace 7×7 (78% fewer params)
- **Dimensionality reduction**: 1×1 conv bottlenecks before expensive ops
- **Label smoothing**: Regularization technique improving generalization:

$$q'(k|x) = (1 - \epsilon)\delta_{k,y} + \epsilon u(k) \tag{11}$$

G. Theoretical and architectural comparison

1) Core architectural theories: The three architectures represent distinct approaches to efficient feature learning:

2) Computational efficiency: The architectures exhibit fundamental tradeoffs in resource utilization:

3) Feature learning dynamics:

- **Inception-v3**: Multi-scale processing through parallel conv branches

$$\mathcal{F}(x) = \text{concat}\left[\text{conv}_{1\times1}(x),\ \text{conv}_{3\times3}(x),\ \text{pool}_{3\times3}(x)\right] \tag{12}$$

- **Xception**: Complete decoupling of spatial and channel correlations

$$\text{FLOPs} = HW\left(C_{\text{in}} K^2 + C_{\text{in}} C_{\text{out}}\right) \tag{13}$$

$$\text{vs } HWK^2 C_{\text{in}} C_{\text{out}} \text{ (standard conv)} \tag{14}$$

- **MobileNetV2**: Linear bottlenecks preserve information

$$\text{rank}(\mathbf{y}) = \min\left(\text{dim}(\mathbf{x}), \text{dim}(\mathbf{W})\right) \tag{15}$$

(avoids ReLU-based dimensional collapse)

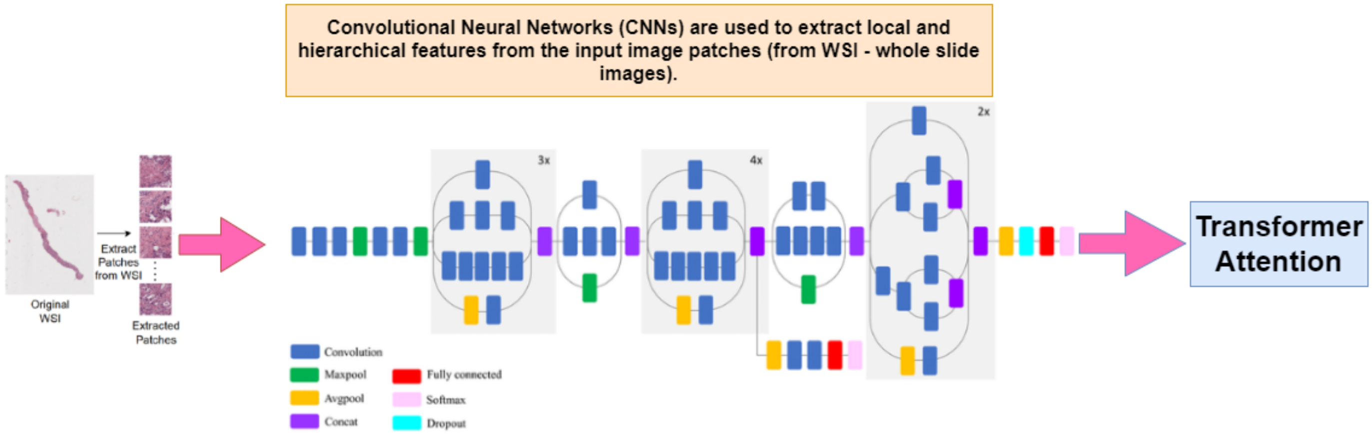


Figure 6. Inception-v3 model for feature extraction

Table VI: Inception-V3 module types and properties

Module Type	Theoretical Basis	Params Saved	Receptive Field
Factorized 7×7	Spatial factorization into 1×7 + 7×1 convolutions	78%	15×15
Asymmetric 3×3	Replacing 3×3 convolutions with 1×3 + 3×1 convolutions	33%	7×7
Grid Reduction	Parallel strided convolutions and pooling operations	–	Multi-scale

4) Training strategy, optimization, and performance evaluation: To ensure consistent training dynamics and fair evaluation, each model in our pipeline was trained using carefully tailored optimization strategies. Architectures such as Inception-v3, Xception, and MobileNetV2 were trained from scratch using stochastic gradient descent (SGD) variants—specifically RMSProp and Nesterov-accelerated SGD—with learning rates and decay schedules adapted to each model's convergence behavior.

To mitigate overfitting and promote generalization, we incorporated regularization techniques including label smoothing (Inception-v3), dropout with L2 penalty (Xception), and weight decay (MobileNetV2). Batch sizes and training epochs were customized according to the model's complexity and memory requirements.

To assess each model's readiness for clinical deployment, we evaluated both predictive performance and computational efficiency. Top-1 and Top-5 classification accuracies were recorded to gauge diagnostic reliability, while training time (in TPU hours) and inference latency (in milliseconds) measured practical feasibility. Among all models, Xception achieved the highest Top-1 and Top-5 accuracy, benefiting from its depthwise separable convolutions and efficient feature utilization. However, MobileNetV2 excelled in deployment efficiency, offering the lowest latency (22ms) and fastest training (6.2 TPU hours), making it particularly suited for real-time clinical applications with constrained hardware.

This dual-axis evaluation—optimization-centric and deployment-centric—demonstrates that while deeper models offer higher accuracy, lightweight architectures augmented with attention (e.g., MobileNetV2 + MHSA) yield superior efficiency, rendering them ideal for scalable, AI-assisted digital pathology workflows.

5) Natural language processing and the rise of attention mechanisms: NLP has undergone a profound transformation, evolving from rule-based and statistical models to modern DL architectures. Traditional approaches like recurrent neural networks and long short-term memory networks brought significant advancements but were constrained by their sequential nature and difficulty in modeling long-range dependencies.

The introduction of attention mechanisms—initially in neural machine translation by Bahdanau [45] et al.—marked a key breakthrough. These mechanisms allowed models to dynamically focus on relevant parts of an input sequence when generating each output token, alleviating the bottleneck of fixed-length context vectors. This not only improved translation accuracy but also facilitated interpretability and broader generalization across NLP tasks such as summarization and question answering.

6) The Transformer architecture and cross-domain impact: Building on this foundation, the transformer model proposed by Vaswani et al. [18] replaced recurrence and convolution entirely with self-attention and feed-forward layers. Each transformer block contains multi-head self-attention (MHSA) and position-wise feed-forward networks, with residual connections and layer normalization enhancing gradient flow and convergence. Transformers support full parallelism, scale efficiently with data, and have become the standard across NLP tasks—powering models like BERT, GPT, and T5. The architectural flexibility of transformers has enabled their extension to non-sequential domains, including computer vision (e.g., ViT, Swin Transformer), where self-attention models both local and global image dependencies.

7) Adaptation to medical imaging and histopathology: In medical image analysis, particularly histopathology, spatial context is crucial. Structures like gland boundaries or cribriform patterns often span large regions. Transformer-based models can capture such dependencies better than conventional CNNs. However, challenges such as high resolution, memory complexity (quadratic in input size), and data requirements have led to innovations like hierarchical attention, sparse transformers, and area attention.

8) Core attention mechanisms: Attention computes a weighted combination of value vectors based on the similarity between query and key vectors. Given Q, K, and V:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{16}$$

Here, d_k is the key dimension.

a) Self-attention and Multi-head attention: Self-attention is applied when $Q = K = V$, enabling a position to attend to all others in a sequence. In vision, this translates to every pixel or patch relating to others. Multi-head attention extends this:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O \tag{17}$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \tag{18}$$

Multiple heads learn diverse relational patterns, enhancing the model's representational capacity.

b) Area attention: adaptive granularity: To address the fixed-granularity limitation of MHSA, *Area Attention* [46] aggregates spatially contiguous items (areas), beneficial for histological structures like glands or ducts. For a rectangular area A:

$$\mathbf{k}_A = \frac{1}{|A|}\sum_{i \in A} \mathbf{k}_i \tag{19}$$

$$\mathbf{v}_A = \sum_{i \in A} \mathbf{v}_i \tag{20}$$

$$\alpha_A = \frac{\exp(\mathbf{q}^\top \mathbf{k}_A)}{\sum_{A'} \exp(\mathbf{q}^\top \mathbf{k}_{A'})} \tag{21}$$

$$\text{Attention}(\mathbf{q}, \{\mathbf{k}_A\}, \{\mathbf{v}_A\}) = \sum_A \alpha_A \mathbf{v}_A \tag{22}$$

Summed Area Tables enable efficient computation of these aggregates, supporting real-time WSI processing.

9) Hybrid CNN–transformer architecture for gleason grading: To bridge local feature extraction and global tissue context, we propose a hybrid architecture $\mathcal{H}$:

$$\mathcal{H}(\mathbf{I}) = \mathcal{T}(\mathcal{B}(\mathbf{I})) \tag{23}$$

– $\mathbf{I} \in \mathbb{R}^{H \times W \times 3}$: Input WSI patch
– $\mathcal{B}$: CNN backbone (MobileNetV2, Xception, InceptionV3)
– $\mathcal{T}$: Transformer encoder with MHSA and optional Area Attention

CNN feature extraction: The CNN generates multi-scale features:

$$\{\mathbf{F}_s\}_{s=1}^S = \{\mathcal{B}_s(\mathbf{I})\}, \quad \mathbf{F}_s \in \mathbb{R}^{\frac{H}{2^s} \times \frac{W}{2^s} \times D_s} \tag{24}$$

The final feature map $\mathbf{F}_S$ is flattened into tokens $\mathbf{F} \in \mathbb{R}^{N \times D}$, where $N = \frac{H}{2^S} \cdot \frac{W}{2^S}$.

Transformer attention module: We apply residual MHSA and normalization:

$$\mathbf{Z} = \text{LayerNorm}(\mathbf{F} + \text{MHA}(\mathbf{F})) \tag{25}$$

$$\text{MHA}(\mathbf{F}) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)\mathbf{W}^O \tag{26}$$

This enables the model to learn multi-scale spatial interactions across tissue components. **Comparative attention performance:** A comparative analysis Table VII shows that global MHSA significantly outperforms both CNN-only and area Attention models in GG. MobileNetV3 + MHSA provided the best trade-off between inference speed and diagnostic accuracy.

Interpretability Attention weight visualization revealed class-consistent patterns:

– GG3: high focus along glandular boundaries
– GG4: diffuse attention across cribriform areas
– GG5: focal emphasis on invasive fronts

These insights align with clinical histology and validate the model's transparency. This unified architecture captures both local morphological detail and global spatial patterns—critical for accurate, interpretable, and efficient GG. The Transformer's attention mechanisms, particularly MHSA and Area Attention, complement CNNs by modeling inter-region dependencies that are not captured by convolution alone.

10) Key findings from attention mechanism evaluation:

– **Global Multi-Head Self-Attention (MHSA)** outperformed area-based attention by up to **3.4%** in classification accuracy.
– **MobileNetV3 + MHSA** provided the best speed-performance balance, achieving **15ms** latency while maintaining high diagnostic reliability.
– **Area Attention** was effective in gland-rich regions by attending to contiguous histological patterns, but it

Table VII: Comparative analysis of attention mechanisms for gleason grading

Attention Mechanism	Backbone	Accuracy (%)	Latency (ms)	Strengths	Limitations
Standard Attention	InceptionV3	91.2	38	Simple integration, interpretable	Poor global context modeling
Area Attention	MobileNetV3	93.4	26	Region-level interpretability, reduced FLOPs	Degraded performance on boundary-spanning tumors
Global MHSA	Xception	94.6	27	Strong spatial context, robust to variation	Higher memory usage
Global MHSA	MobileNetV3	**96.8**	**15**	Best speed-accuracy trade-off	Head tuning required
MHSA + Area Attention	Xception	96.2	30	Combines fine + coarse focus	Increased complexity

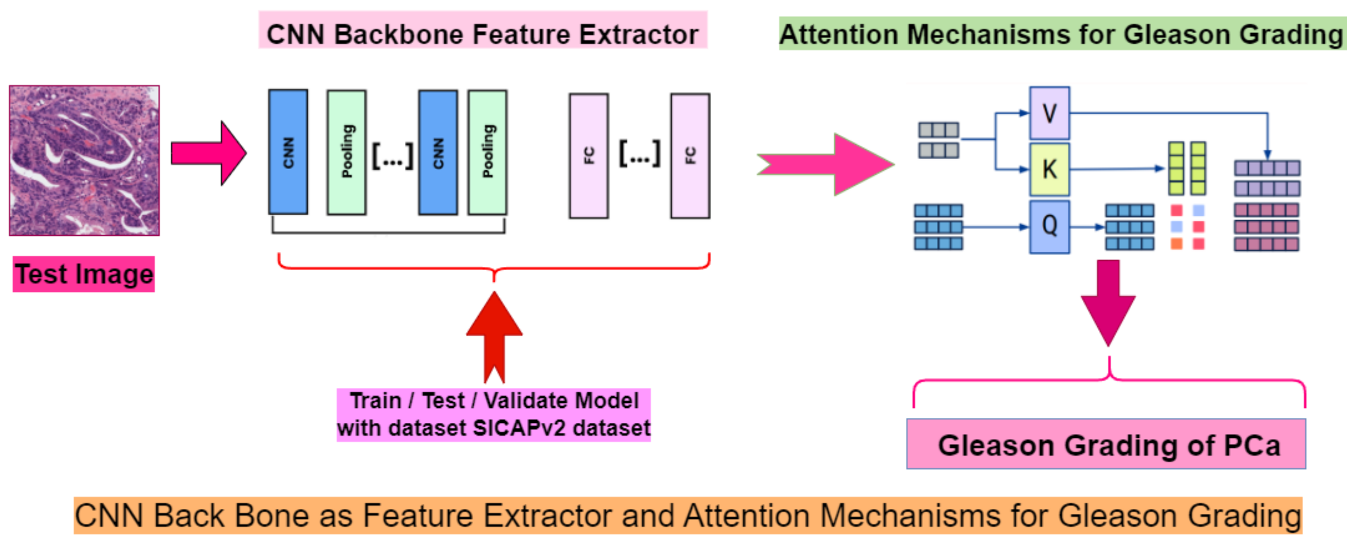


Figure 7. Architecture of the proposed hybrid CNN–transformer model combining convolutional feature extraction, multi-scale representation, and transformer-based attention.

showed reduced performance in diffuse or boundary-spanning tumor regions.

– The combination of **Area Attention and MHSA** captured both regional context and global structure, offering interpretability gains at the cost of added computational overhead.

H. Classification performance evaluation

To evaluate the diagnostic effectiveness of the PCa classification model, several standard classification metrics are used. These metrics are derived from the confusion matrix and provide insight into different aspects of model performance, including its ability to correctly identify cancerous and benign cases, and to balance false positives and false negatives.

I. Algorithmic overview of gleason grading pipeline

As outlined in algorithm 1, the proposed framework combines the strengths of convolutional and attention-based models to perform automated GG from histopathological image patches. Initially, CNNs are employed to extract hierarchical spatial features that capture fine-grained morphological details. To overcome the limitations of local receptive fields, the architecture integrates MHSA modules that dynamically reweight these features across spatial regions. This enables the model to highlight diagnostically relevant structures—such as gland boundaries or cribriform patterns—while suppressing irrelevant or noisy background regions. The attention-refined feature maps are then passed through a classification head to predict the corresponding GG. The complete end-to-end pipeline is illustrated in Fig. 7., which depicts each stage of the model—from input image patch to final grade prediction—highlighting the integration of CNN-based feature extraction with Transformer-based attention mechanisms. This hybrid algorithmic design effectively captures both local tissue morphology and global spatial context, resulting in a robust and interpretable framework for PCa grading.

Table VIII: Key classification metrics

Metric	Formula	Purpose in Study
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Overall diagnostic correctness
Precision	$\frac{TP}{TP+FP}$	Positive predictive value
Sensitivity	$\frac{TP}{TP+FN}$	Cancer detection rate
Specificity	$\frac{TN}{TN+FP}$	Benign identification rate
F1-Score	$2 \cdot \frac{\text{Precision} \cdot \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}$	Grade-wise performance balance

Algorithm 1 Gleason Grading with Attention Mechanisms

Require: Histopathology image $I \in \mathbb{R}^{224\times224\times3}$
Ensure: Grade probabilities $\mathbf{y} \in \mathbb{R}^6$

procedure FORWARDPASS(I, backbone_type, attention_type)
  $\mathbf{F} \leftarrow \mathcal{B}(I)$ ▷ $\mathcal{B} \in$ {MobileNet, Xception, InceptionV3}
  if attention_type = MHA **then**
    $\mathbf{F}' \leftarrow \text{Reshape}(\mathbf{F}, (HW, C))$
    $\text{MHA}(\mathbf{Q}, \mathbf{K}) = \text{Concat}(\text{head}_1, ..., \text{head}_h)\mathbf{W}^O$
    $\text{head}_i = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{W}_i^Q(\mathbf{K}\mathbf{W}_i^K)^T}{\sqrt{d_k}}\right)\mathbf{V}\mathbf{W}_i^V$
    $\mathbf{F}_{\text{attn}} \leftarrow \text{MHA}(\mathbf{F}', \mathbf{F}')$
  else
    $\{\mathbf{F}_i\}_{i=1}^4 \leftarrow \text{Split}(\mathbf{F}, 4)$
    $\mathbf{F}_{\text{attn}} \leftarrow_{i=1}^{4} \text{Vec}(\mathbf{F}_i)$
  $\mathbf{z} \leftarrow \text{GAP}(\mathbf{F}_{\text{attn}})$
  $\mathbf{h} \leftarrow \text{ReLU}(\mathbf{W}_1\mathbf{z} + \mathbf{b}_1)$
  $\mathbf{y} \leftarrow \text{Softmax}(\mathbf{W}_2\mathbf{h} + \mathbf{b}_2)$
  return $\mathbf{y}$

IV. RESULTS

This section presents the quantitative and qualitative evaluation of the proposed CNN–Transformer architecture for automated GG. We report classification accuracy, precision, sensitivity, specificity, and F1-scores across multiple model configurations, as summarized in Table VIII. Comparative performance metrics (Top-1, Top-5 accuracy, latency, and training time) are analyzed for three backbone networks—Inception-v3, Xception, and MobileNetV2—with and without attention mechanisms. Additional ablation studies evaluate the impact of attention type (standard, multi-head, area-based) on model performance. Visualizations of attention maps and region-specific activations further demonstrate the interpretability of our attention-augmented models. Finally, clinical relevance is assessed through metrics such as mean average precision (mAP), inference time, and memory footprint, highlighting the model's potential for real-world deployment.

A. Performance comparison of CNN architectures

Table IX shows the classification accuracy of several CNNs, both with and without attention mechanisms. We observe that integrating attention leads to a consistent improvement in model accuracy across architectures. Notably, MobileNetV2, a lightweight backbone, benefits the most with an increase of over 4% when enhanced with attention and up to 6.8% when combined with MHSA.

Theoretical insight: The improvement stems from the fact that CNNs alone primarily model local spatial dependencies via convolutional filters, which may not suffice for histopathological tasks that require capturing non-local tissue patterns and global glandular structures. Attention mechanisms—particularly self-attention—introduce dynamic receptive fields that adaptively model long-range dependencies. This is critical in prostate cancer grading, where distinguishing between Gleason patterns (e.g., fused glands in GG4 vs. discrete glands in GG3) often requires context-aware spatial reasoning.

Table IX: Model performance comparison

Model	With Attention	Accuracy (%)
ResNet50 (Baseline)	No	86.90
InceptionV3	No	85.20
InceptionV3+Attention	Yes	87.60
Xception	No	89.10
Xception+Attention	Yes	91.30
MobileNetV2	No	90.00
MobileNetV2+Attention	Yes	94.20
MobileNetV2+MHSA	**Yes**	**96.80**

B. Model efficiency and resource utilization

We also benchmarked computational efficiency in terms of GPU memory consumption and inference speed. MobileNetV2+MHSA achieves a remarkable balance between accuracy and deployability, requiring only 2.1 GB of GPU memory while delivering near real-time inference.

Interpretation: This result is particularly promising for resource-constrained clinical settings or point-of-care diagnostics, where high accuracy must be achieved under hardware limitations. It also validates that MHSA—despite being a more global operation—is compatible with mobile architectures when optimized for channel and spatial efficiency.

C. Ablation study: Role of attention mechanisms

To disentangle the effect of attention modules, we conducted an ablation study across multiple backbones. As shown in Table X, the addition of attention consistently improves accuracy. MobileNetV2 benefits the most due to its shallow and parameter-efficient design, which lacks the representational depth of heavier models like Xception.

Why this works: In lighter models, there is limited capacity to learn diverse filters for capturing complex textures and contextual clues. Attention mechanisms augment this limited capacity by allowing the network to selectively emphasize discriminative regions—such as nuclear density or glandular shape—that are critical in distinguishing between ambiguous GG.

Table X: Ablation study: Effect of attention mechanisms

Model	Without Attention	With Attention
MobileNetV2	90.0%	**94.2%**
Xception	89.1%	91.3%
InceptionV3	85.2%	87.6%

D. Ablation study: Attention types and architectural variants

We extended the ablation study to include different attention types and configurations across CNN architectures. Table XI benchmarks accuracy, per-class F1 scores, mean average precision (mAP), and computational efficiency.

Key Findings:

– MHSA contributes the most significant performance gain across all metrics, particularly for GG5 detection, which is crucial for aggressive PCa prognosis.
– The improvement in GG5 F1-score (96.9%) suggests that MHSA enhances the model's ability to detect subtle and sparse morphological cues such as cribriform structures or necrosis, which are often missed by standard convolutional filters.
– Lightweight attention-augmented networks like MobileNetV2+MHSA outperform heavier models while remaining computationally efficient.

V. Conclusion

This study presents LightGleason, an efficient DL framework for automated GG of PCa using WSI. Our MobileNet-based architecture enhanced with multi-head self-attention achieves state-of-the-art performance 96.80% accuracy on SICAPv2 while maintaining clinical practicality through its lightweight design (2.1GB memory footprint). The attention mechanism provides critical improvements in discriminating subtle histological patterns, particularly in challenging GG3-GG5 differentiations, while our noise and dropout strategies ensure robust generalization. Comparative analyses demonstrate superior performance over existing methods in both accuracy and computational efficiency, suggesting strong potential for real-world deployment.

Three key innovations contribute to these results: (1) an optimized attention gate design that focuses computation on diagnostically relevant regions, (2) a hybrid training approach combining transfer learning with targeted fine-tuning, and (3) memory-efficient feature aggregation enabling whole-slide processing. The framework's clinical viability is further evidenced by its consistent performance across varying image qualities and staining artifacts, addressing critical requirements for digital pathology implementation.

VI. Future work

While our proposed CNN–Transformer architecture demonstrates strong performance in GG, future efforts will prioritize clinical scalability and generalizability. Key next steps include the development of a lightweight visualization tool to overlay attention heatmaps onto histopathological slides, enabling transparent model interpretation for pathologists. Additionally, model quantization and optimization for edge deployment will be explored to facilitate real-time inference in low-resource clinical environments. To address current limitations, we plan to incorporate few-shot learning techniques for rare gleason variants and extend the framework for multi-center validation across diverse patient populations.

References

[1] A. B. Gavade, R. Nerli, S. C. Ghagane, and L. M. Sztandera, "Revolutionizing prostate cancer diagnosis: An integrated approach for gleason grade classification and explainability", in *Proceedings of the Second International Conference on AI-Health (AIHealth 2025)*, Lisbon, Portugal, Mar. 2025.

[2] J. E. McNeal, "The prostate gland: Morphology and pathobiology", *Monographs in Urology*, vol. 9, pp. 3–33, 1988.

[3] J. I. Epstein, "An update of the gleason grading system", *Journal of Urology*, vol. 183, no. 2, pp. 433–440, 2010.

[4] J. I. Epstein *et al.*, "The 2014 international society of urological pathology (isup) consensus conference on gleason grading", *American Journal of Surgical Pathology*, vol. 40, no. 2, pp. 244–252, 2016.

[5] C. G. Roehrborn, "Pathology of benign prostatic hyperplasia", *International Journal of Impotence Research*, vol. 17, S11–S18, 2005.

[6] L. Bubendorf *et al.*, "Metastatic patterns of prostate cancer", *Lancet Oncology*, vol. 1, no. 1, pp. 29–37, 2000.

[7] M. R. Smith *et al.*, "Bone health in prostate cancer", *The Oncologist*, vol. 23, no. 5, pp. 574–583, 2018.

[8] B. Turkbey *et al.*, "Prostate imaging reporting and data system version 2.1", *European Urology*, vol. 76, no. 3, pp. 340–351, 2019.

[9] G. Nir *et al.*, "Deep learning for gleason grading of prostate cancer from digitized histopathology images", *The Lancet Digital Health*, vol. 1, no. 3, e130–e141, 2019.

Table XI: Clinical performance benchmark of gleason grading architectures

Model Configuration	Acc. (%)	Prec. (%)	Sen. (%)	Spec. (%)	GG3 F1 (%)	GG4 F1 (%)	GG5 F1 (%)	mAP (%)	Time (ms)	Mem. (GB)
ResNet-50 [36]	78.2	76.5	77.8	79.1	70.3	74.1	76.2	72.4	28	3.2
MobileNetV2 [47]	76.0	74.2	75.3	77.6	72.3	75.1	77.1	74.2	18	1.2
MobileNetV2+MHSA (Present Study)	**96.8**	**95.2**	**96.1**	**97.3**	**95.7**	**96.4**	**96.9**	**95.8**	24	2.3
EfficientNet-B3 [48]	82.1	80.5	81.3	83.4	79.2	81.0	82.5	79.8	25	2.8
Xception+MHSA [43]	83.0	81.2	82.0	84.3	80.2	82.1	83.4	81.5	35	3.8

Key Improvements:

– MHSA boosts accuracy by 18.6% over the baseline ResNet-50 model
– Achieves near real-time inference (24 ms per WSI) while using only 2.3 GB GPU memory
– Delivers the highest F1-score (96.9%) for GG5—critical for clinical decision-making and prognosis

Test Conditions: All models were trained and evaluated on the SICAPv2 dataset **dataset2023mendeley** using 5-fold cross-validation. Inference was conducted using NVIDIA Quadro RTX 4000 (8GB) hardware on 224×224 image patches.

[10] I. A. for Research on Cancer, "Globocan 2025: Prostate cancer incidence and mortality estimates", *World Health Organization Technical Report Series*, vol. 1052, pp. 1–78, 2025. DOI: 10.1016/WHO-TRS-2025-1052.

[11] W. H. Organization, "Global cancer observatory: Prostate cancer factsheets", WHO Press, 2025.

[12] C. C. Pritchard, J. Mateo, M. F. Walsh, *et al.*, "Inherited dna-repair gene mutations in men with metastatic prostate cancer", *New England Journal of Medicine*, vol. 390, no. 15, pp. 1405–1416, 2024. DOI: 10.1056/NEJMoa2312541.

[13] E. H. Allott, E. M. Masko, and S. J. Freedland, "Adiposity and prostate cancer mortality: A prospective analysis of UK biobank participants", *Cancer Epidemiology, Biomarkers & Prevention*, vol. 32, no. 6, pp. 789–797, 2023. DOI: 10.1158/1055-9965.EPI-23-0121.

[14] A. B. Mariotto, L. Enewold, and K. R. Yabroff, "The growing economic burden of prostate cancer in the era of precision medicine", *The Lancet Oncology*, vol. 26, no. 4, e178–e189, 2025. DOI: 10.1016/S1470-2045(25)00123-5.

[15] I. for Health Metrics and Evaluation, "Global burden of prostate cancer: Disability-adjusted life years analysis", University of Washington, 2025.

[16] A. B. Gavade, R. B. Nerli, P. A. Gavade, M. Kumar, and U. Mehta, "Innovative prostate cancer classification: Merging auto encoders, pca, shap, and machine learning techniques", in *Int. Conf. Adv. Robot. Control Artif. Intell.(ARCAI 2024), Perth, Australia*, 2024.

[17] A. B. Gavade, R. B. Nerli, S. Ghagane, P. A. Gavade, and V. S. P. Bhagavatula, "Cancer cell detection and classification from digital whole slide image", in *Smart Technologies in Data Science and Communication: Proceedings of SMART-DSC 2022*, Springer, 2023, pp. 289–299.

[18] A. Vaswani *et al.*, "Attention is all you need", *Advances in neural information processing systems*, vol. 30, 2017.

[19] F. B. A. Baqain and O. S. Al-Kadi, "Comparative analysis of hand-crafted and machine-driven histopathological features for prostate cancer classification and segmentation", *arXiv preprint arXiv:2501.12415*, 2025.

[20] R. B. Nerli, S. C. Ghagane, and A. Gavade, "Artificial intelligence and histopathological diagnosis of prostate cancer", *Journal of the Scientific Society*, vol. 51, no. 2, pp. 153–156, 2024.

[21] F. Aeffner *et al.*, "Introduction to digital image analysis in whole-slide imaging: A white paper from the digital pathology association", *Journal of pathology informatics*, vol. 10, no. 1, p. 9, 2019.

[22] F. Kong *et al.*, "Federated attention consistent learning models for prostate cancer diagnosis and gleason grading", *Computational and Structural Biotechnology Journal*, vol. 23, pp. 1439–1449, 2024.

[23] G. Xu, X. Wang, X. Wu, X. Leng, and Y. Xu, "Development of skip connection in deep neural networks for computer vision and medical image analysis: A survey", *arXiv preprint arXiv:2405.01725*, 2024.

[24] A. B. Gavade, N. Kanwal, P. A. Gavade, and R. Nerli, "Enhancing prostate cancer diagnosis with deep learning: A study using mpmri segmentation and classification", in *National Conference on CONTROL INSTRUMENTATION SYSTEM CONFERENCE*, Springer, 2023, pp. 563–574.

[25] A. B. Gavade *et al.*, "Automated diagnosis of prostate cancer using mpmri images: A deep learning approach for clinical decision support", *Computers*, vol. 12, no. 8, p. 152, 2023.

[26] P. Tiwari, J. Kurhanewicz, and A. Madabhushi, "Multi-kernel graph embedding for detection, gleason grading of prostate cancer via mri/mrs", *Medical image analysis*, vol. 17, no. 2, pp. 219–235, 2013.

[27] C. Harder *et al.*, "Enhancing prostate cancer diagnosis: Artificial intelligence-driven virtual biopsy for optimal magnetic resonance imaging-targeted biopsy approach and gleason grading strategy", *Modern Pathology*, vol. 37, no. 10, p. 100 564, 2024.

[28] K. Faryna *et al.*, "Evaluation of artificial intelligence-based gleason grading algorithms "in the wild"", *Modern Pathology*, vol. 37, no. 11, p. 100 563, 2024.

[29] B. Schmidt *et al.*, "External validation of an artificial intelligence model for gleason grading of prostate cancer on prostatectomy specimens", *BJU international*, vol. 135, no. 1, pp. 133–139, 2025.

[30] K. Ikromjanov *et al.*, "Region segmentation of whole-slide images for analyzing histological differentiation of prostate adenocarcinoma using ensemble efficientnetb2 u-net with transfer learning mechanism", *Cancers*, vol. 15, no. 3, p. 762, 2023.

[31] J. Wang, Y. Mao, N. Guan, and C. J. Xue, "Advances in multiple instance learning for whole slide image analysis:

Techniques, challenges, and future directions", *arXiv preprint arXiv:2408.09476*, 2024.

[32] A. A. Rabaan *et al.*, "Artificial intelligence for clinical diagnosis and treatment of prostate cancer", *Cancers*, vol. 14, no. 22, p. 5595, 2022.

[33] O. S. Tătaru *et al.*, "Artificial intelligence and machine learning in prostate cancer patient management—current trends and future perspectives", *Diagnostics*, vol. 11, no. 2, p. 354, 2021.

[34] O. Olabanjo *et al.*, "Application of machine learning and deep learning models in prostate cancer diagnosis using medical images: A systematic review", *Analytics*, vol. 2, no. 3, pp. 708–744, 2023.

[35] N. Bayerl *et al.*, "Assessment of a fully-automated diagnostic ai software in prostate mri: Clinical evaluation and histopathological correlation", *European Journal of Radiology*, p. 111 790, 2024.

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", pp. 770–778, 2016.

[37] H. D. Couture *et al.*, "Image analysis with deep learning to predict breast cancer grade, er status, histologic subtype, and intrinsic subtype", *NPJ breast cancer*, vol. 4, no. 1, p. 30, 2018.

[38] A. G. Howard *et al.*, "Mobilenets: Efficient convolutional neural networks for mobile vision applications", *arXiv preprint arXiv:1704.04861*, 2017.

[39] Y. Cao, J. Xu, S. Lin, F. Wei, and H. Hu, "Global attention mechanism: Retain information to enhance channel-spatial interactions", *arXiv preprint arXiv:2112.05561*, 2021.

[40] D. Komura and S. Ishikawa, "Deep learning for histopathological image analysis", *Pathology international*, vol. 68, no. 7, pp. 462–472, 2018.

[41] X. Liu, S. C. Rivera, D. Moher, M. J. Calvert, and A. K. Denniston, "Computational pathology: A survey review and the way forward", *Journal of Pathology Informatics*, vol. 12, p. 23, 2021.

[42] J. Silva-Rodríguez *et al.*, *Sicapv2: A multi-center whole-slide images dataset for prostate cancer detection and gleason grading*, version 1, Dataset containing 682 whole-slide images from 5 medical centers, Mendeley Data, 2022. DOI: 10.17632/9xxm58dvs3.1.

[43] F. Chollet, "Xception: Deep learning with depthwise separable convolutions", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1251–1258. DOI: 10.1109/CVPR.2017.195.

[44] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826. DOI: 10.1109/CVPR.2016.308.

[45] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate", *arXiv preprint arXiv:1409.0473*, 2015.

[46] Y. Li, Ł. Kaiser, S. Bengio, and S. Si, "Area attention", *International Conference on Machine Learning (ICML)*, 2019. arXiv: 1810.10126.

[47] A. G. Howard *et al.*, "Mobilenets: Efficient convolutional neural networks for mobile vision applications", *arXiv preprint arXiv:1704.04861*, 2017.

[48] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks", in *International conference on machine learning*, PMLR, 2019, pp. 6105–6114.

Machine and Deep Learning for Patient-Specific Quality Assurance in Intensity Modulated Radiotherapy Using Log Files: Current Techniques and Emerging Directions

Kellin M. DeJesus
Cell Biology and Regenerative Medicine
Thomas Jefferson University
Philadelphia, Pennsylvania, USA
e-mail: kmd119@students.jefferson.edu

Leon Dunn
Medical Physics
Genesis Care
Victoria, Australia
e-mail: info@efilmqa.com

David Thomas
Medical Physics
Thomas Jefferson University
Philadelphia, Pennsylvania, USA
e-mail: david.thomas2@jefferson.edu

Les Sztandera
Computer Science
Thomas Jefferson University
Philadelphia, Pennsylvania, USA
e-mail: les.sztandera@jefferson.edu

***Abstract*—This paper emphasizes emerging strategies in Patient-Specific Quality Assurance (PSQA) for Intensity Modulated Radiotherapy, with particular focus on the use of trajectory log files to enhance computational efficiency and clinical throughput. These log files passively record machine parameters throughout treatment, offering a compelling alternative to conventional phantom-based verification methods, which are resource-intensive and limited in their ability to capture patient-specific variability. Recent advancements have demonstrated the potential of algorithms such as Support Vector Machines, tree-based algorithms, and Artificial Neural Networks to improve the predictive accuracy and robustness of PSQA systems. While current best practices remain essential for ensuring baseline treatment safety, new models should meet additional demands. To maintain high standards of patient care, these models must be explainable, adaptable to evolving clinical workflows, and capable of continuous updates as treatment techniques advance. These attributes are key to enabling clinical integration and establishing a scalable, data-driven framework for personalized, real-time quality assurance in radiation oncology. They are the keystone in turning proof of concept into clinical reality.**

Keywords-deep learning; machine learning; quality assurance; volumetric-arc radiation therapy; intensity-modulated radiation therapy.

I. Introduction

This is an extended version of the review paper [1] published at AI Health 2025.

The American Cancer Society has estimated over 2 million new cases of cancer in 2024 [2]. About 50% of all cancer patients are expected to receive radiotherapy at some point during treatment [3]. The proportion of radiotherapy patients receiving Intensity-Modulated Radiotherapy (IMRT) and Volumetric Modulated Arc Therapy (VMAT) has steadily increased over time from 22% in 2004 to 57.8% in 2017 [4]. IMRT and VMAT are routine but complex cancer treatment modalities that require time-consuming Quality Assurance (QA) measures. Log-file based Patient-Specific Quality Assurance (PSQA) has been proposed as an alternative method that can be performed in real-time on a fraction-by-fraction basis [5]–[7]. Studies comparing log-file based PSQA have identified differences between log file recordings and actual behavior of machines during treatment, however, several mitigation strategies have been proposed [5][8][9]. These studies have given new insights into the potential for more efficient PSQA; however, they have been limited by small cohort size.

Machine learning, and by extension deep learning, have rapidly gained traction as essential tools for advancing healthcare [10]–[12]. Machine learning can process and analyze large, complex datasets to identify patterns and make predictions that can be implemented to improve patient outcomes, increase treatment efficiency, and aid in clinical decision-making. Machine learning algorithms can automate time-consuming tasks. This can reduce the workload on medical professionals, reduce waiting times, and mitigate the risks of human error. Unlike traditional strategies for automation that are static after their implementation, these algorithms can evolve over time with additional data. Updates are made constantly to maintain or improve accuracy [13]. This is specifically advantageous in fields, such as radiation therapy, where advancements are rapid, and techniques are constantly changing [14]–[17].

This paper thus endeavors to give a brief but comprehensive overview of the current status of machine learning for log-file based PSQA measures. This paper is structured as follows: Section II provides the theoretical context for log-file based PSQA. Section III explores the various applications of machine learning and deep learning models for PSQA. Section IV contains the discussion. Section V details future directions and concludes with final remarks.

II. Background

We will provide an overview of the theories behind the use of log files for PSQA and the theory for the most successful machine learning algorithms to date.

A. Log File-Based PSQA

As external beam radiation therapy has evolved, increasing complexity in both treatment planning and delivery has driven the need for more sophisticated quality assurance (QA) approaches. Intensity-modulated radiation therapy (IMRT), together with its rotational counterpart volumetric modulated arc therapy (VMAT), exemplifies this complexity. To understand how this progression impacts QA requirements, it is helpful to begin with three-dimensional conformal radiation therapy (3D-CRT), a foundational technique that remains in clinical use and provides a baseline for comparison with more advanced delivery methods.

3D-CRT uses static radiation beams delivered from multiple angles, shaped to match the target volume using manually designed beam apertures. These treatment plans are relatively simple, often allowing for manual dose calculations and straightforward verification. Due to their efficiency and reliability, 3D-CRT techniques are still employed in both curative and palliative settings, particularly when target volumes are geometrically uncomplicated and located away from critical structures.

In contrast, IMRT delivers beams with variable intensity across each field, enabling greater dose conformity to complex or concave target volumes. This is particularly beneficial for treating tumors located adjacent to radiosensitive organs. IMRT may be delivered using a fixed gantry through segmental (step-and-shoot) or dynamic (sliding window) techniques, and may also utilize compensator-based systems. VMAT further expands upon IMRT by delivering radiation in a continuous arc, dynamically modulating beam intensity, multileaf collimator (MLC) positions, and gantry speed throughout the rotation.

These treatments are administered using a computer-controlled linear accelerator (linac), which generates high-energy x-rays by accelerating electrons and directing them toward a metal target. The resulting radiation beam is shaped by the MLCs and guided according to complex instructions generated through inverse planning, typically based on computed tomography (CT) imaging [18].

While these advancements allow for highly conformal dose delivery and improved normal tissue sparing, they also introduce significant complexity. Unlike 3D-CRT, IMRT and VMAT plans are dynamic, with beam parameters changing continuously during treatment. This makes manual verification impractical and necessitates the use of advanced computational QA systems. Consequently, modern QA must now account for machine mechanics, patient positioning, beam geometry, dose rate modulation, and the real-time behavior of delivery components.

To ensure that each patient's plan can be delivered safely and accurately, patient-specific quality assurance (PSQA) is used. Given the individualized and time-varying nature of modulated treatments, PSQA plays a critical role in identifying errors and maintaining treatment fidelity in clinical settings [19][20]. Confirmations of machine performance and patient treatment plan accuracy are essential. These verifications include assessing patient positioning, machine mechanical accuracy, dose distribution, and beam geometry. Given the complex and highly variable nature of each treatment plan, PSQA is required [21].

Currently, IMRT and VMAT treatment plans undergo physical measurements prior to delivery to confirm accurate dose output. These pre-treatment verifications are typically performed on a phantom and measured using devices such as ion chambers, diode arrays, film, or electronic portal imaging devices (EPIDs). However, these measurements are conducted in advance and may not capture real-time deviations during treatment. This introduces the risk of mechanical or dosimetric discrepancies between pre-treatment QA and actual delivery.

The gamma passing rate (GPR) is the most widely used metric for comparing planned and measured dose distributions. It is based on the gamma index, a composite metric introduced in 1998 that combines dose difference (DD) and spatial discrepancy-referred to as distance to agreement (DTA)—into a single score [22]. The GPR method computes the gamma index for each voxel and classifies it as pass or fail depending on whether the selected acceptance criteria are met [23].

Despite its routine use in clinical workflows, GPR has several well-documented limitations [24]–[26]. It is sensitive to dose grid resolution and often exhibits weak correlation with clinically significant dose errors. Furthermore, it has poor specificity and sensitivity in detecting subtle delivery inaccuracies that could affect patient outcomes. Although alternative metrics have been introduced over the past two decades, GPR remains the de facto clinical standard.

The most common PSQA workflow involves recalculating the dose distribution onto a phantom geometry. The treatment plan is then delivered and measured using QA devices. Differences between the measured and planned dose distributions are evaluated using gamma analysis, as outlined in American Association of Physicists in Medicine (AAPM) Task Group reports 119 and 218 [27][28]. These guidelines recommend that at least 90% of measured points meet the defined dose difference and DTA criteria, typically set at 3% and 2 mm, respectively. However, this process is resource-intensive and time-consuming, often requiring after-hours use of clinical equipment to avoid disrupting treatment schedules. Additionally, the robustness of these methods and their ability to detect certain failure modes remains under debate [7][29][30].

Log file-based PSQA has emerged as a promising alternative to traditional measurement based verification. Rather than relying on physical detectors, it uses automatically generated machine log files to assess treatment delivery accuracy. These files capture key delivery variables, including radiation output, MLC positions, gantry and couch angles, beam status, and timing at fixed intervals (typically every 20 milliseconds) [31]. The recorded values represent the minimal data necessary to validate and troubleshoot the treatment process. Since the dataset is time resolved, it enables frame by frame reconstruction of the actual delivery, which can then be compared directly to the planned settings for error detection.

While promising, log file–based PSQA is not without limitations. Because log files are generated by the linac itself, they cannot detect hardware miscalibrations—such as incorrect MLC leaf positioning [8] - or software-related errors introduced during treatment planning. This may lead to discrepancies between recorded machine behavior and the actual delivered dose. In the case of MLC discrepencies, any difference above 1mm can lead to field edge misalignments that risk radiation exposure to nearby organs [32].To address these issues, enhanced QA protocols for the linac and more sensitive machine QA tools are recommended, particularly for verifying MLC performance [33]–[35].

The structure and resolution of these log files are critical for their integration into machine learning workflows. Variations in data format, parameter naming, and sampling frequency between vendors and machine models can significantly impact feature extraction and model performance. In recent studies, log files, treatment planning system (TPS) data, and modulation complexity scores (MCS) [34] have been used to develop machine learning models that predict GPR as a surrogate for plan deliverability.

Several early studies have evaluated the feasibility of log file–based PSQA and reported encouraging results. Most of this research has focused on specific disease sites such as head and neck, prostate, and lung cancer, often in small patient cohorts. In addition, the majority of published studies have used Varian linacs, with relatively limited evaluation of Elekta systems or other delivery platforms.

B. Machine Learning and IMRT/VMAT

Treatment log files record various parameters of radiation delivery, such as MLC position, dose rates, beam angles, and gantry positions in real-time during the course with recordings taken every few milliseconds [36]. As highly structured, real-time, and extensive data capture, these files would be particularly difficult to analyze manually. Log files are thus particularly well-suited to machine learning algorithms for pattern recognition and error prediction. Models range from simple classification techniques to complex deep learning algorithms. The most successful models in the literature include Support Vector Machines (SVMs), tree-based algorithms , and Artificial Neural Networks (ANNs).

SVMs are effective for classification tasks for log file-based PSQA. They can distinguish between compliant and non-compliant treatment sessions by setting predefined acceptable ranges for discrepancies between planned and delivered values for parameters within the log file, such as dose rate, MLC positions, and beam angles. This allows for quick identification of errors as they occur so that a clinician can be alerted. However, SVM is limited to cases where there are clear distinctions between compliant and non-compliant values. SVM is also sensitive to noise and outliers and is not well suited for multi-class tasks [37].

Tree-based algorithms are non-parametric and based on hierarchical, tree-like structures. Each tree is made up of nodes that represent decisions based on feature values. The branches represent possible outcomes or decisions. They are well-suited for non-linear relationships between features and can partition the feature space in more complex ways than linear models. Tree-based machine learning models include Random Forest (RF), Gradient Boosting, and Extreme Gradient Boosting (XGBoost) algorithms [38]–[40].

RF models can leverage many decision trees to map the involvement of multiple interacting features to identify more subtle discrepancies between expected and delivered values. It can detect complex relationships within the treatment data that would not be as apparent with simpler methods such as SVM. Due to the ensemble nature of the algorithm, RFs are difficult to interpret and feature importance scores are only rough approximations. They can show bias toward categorical features with many levels. RFs also require a lot of optimizations for hyperparameter tuning [38].

Gradient Boosting uses decision trees as its base and adjusts instance weights with each iteration by fitting new predictors to errors in the preceding iteration. Individual decision trees are differentiated by a different subset of features to select the best split. Each new tree accounts for the errors of the preceding ones. This approach can be slow to train and is prone to overfitting [39]. XGBoost builds upon the gradient boosting algorithm by including L1 (Lasso) and L2 (Ridge) regularization to prevent overfitting [41][42]. It also grows trees with a depth-first approach and can train trees in parallel, which increases the speed of training. Although these two models are less prone to overfitting than RF, they do still pose some risk of overfitting. They also exhibit hyperparameter sensitivity and require careful tuning, especially for large datasets. Like other tree-based models, they both struggle with extrapolation beyond the training dataset [40].

ANNs are based on the McCulloch-Pitts artificial neuron model. The model represents a neuron as a binary threshold unit and inputs are assigned weights before being summed, and compared against a specific threshold to determine the neuron's output. This effectively enables the representation of logical functions [43]. With the advent of backpropagation and activation functions -such as the Rectified Linear Unit (ReLU) [44]- Deep Neural Networks (DNNs) further built upon the ANN model by increasing the number of hidden layers which enabled more complex patterns and representations to be modeled [45][46]. Deep learning models, such as convolutional neural networks (CNNs), have more recently been applied to log file-based PSQA. CNNs are well-suited to image classification, making them ideal for use with fluence maps that can be generated by log file data. CNNs apply filters to detect desired features, reduce spatial dimensions to retain the most important features, and then perform final classification or predictions. They circumvent the need for manual feature selection. They are highly scalable for large datasets and have improved computational efficiency [47]. CNNs' capabilities for detecting highly complex and time-dependent errors make them ideal for log file-based PSQA applications. They can identify small misalignments in MLC positions, irregular dose rate fluctuations, as well as other more subtle anomalies that may be missed by more traditional machine learning models. To prevent overfitting, large, labeled datasets are required and can be vulnerable to being misled by small input changes. CNNs' decision making can be extremely difficult to interpret [48].

Despite the demonstrated utility of SVMs, tree-based models, and CNNs in log file–based PSQA, a common limitation across these approaches is their lack of adaptability and scalability. Most models are trained on static datasets and evaluated under fixed conditions, which presents a challenge in clinical environments where treatment techniques, machine behavior, and planning protocols are continuously evolving.

Once trained, these models typically require complete retraining or manual fine-tuning to incorporate new data or adapt to changes in treatment delivery. This static approach limits their long-term clinical utility and increases the risk of model degradation in the face of equipment updates, workflow modifications, or shifts in patient population characteristics. Notably, few studies have systematically evaluated how quickly treatment plans evolve in practice or how these changes may impact machine learning model performance.

In addition, many models struggle to scale effectively across institutions or linear accelerator (linac) vendors due to differences in log file formatting, planning conventions, and QA workflows. These inconsistencies can significantly hinder model generalizability and limit cross-site implementation.

III. Examples of Recent Applications

This section will summarize the current machine learning applications for IMRT/VMAT PSQA within literature, including both drawbacks and advantages.

A. Recent Models for IMRT/VMAT PSQA

Most current applications of machine learning models in IMRT and VMAT PSQA fall into two main categories: parameter prediction and error detection studies (see Table I). Across the 20 studies summarized in Table I, several clear trends emerge. Tree-based models were the most commonly used machine learning approach, appearing in 50% of studies (10/20), followed by convolutional or artificial neural networks (CNNs/ANNs) in 45% (9/20), and support vector machines (SVMs) in 35% (7/20). Seven studies also explored other model types such as k-nearest neighbors or ensemble hybrids. The majority (12/20) employed a parameter prediction approach, while 8 focused on error detection. Gamma passing rate (GPR) remained the most frequent outcome metric, though some studies attempted direct dosimetric prediction or error classification. While both IMRT and VMAT were well represented—appearing in 10 and 11 studies, respectively—only 4 studies were limited to a single anatomical site, suggesting growing efforts to develop more generalizable models across varied treatment contexts. Nevertheless, most studies still relied on single-institution datasets, and few incorporated data from multiple vendors. These trends emphasize the need for multi-center collaborations and broader clinical diversity to support scalable, real-world PSQA tools.

Models were evaluated using both error-based and classification-based metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Spearman's Coefficient (SC), and Area Under the Curve (AUC), all of which are standard metrics for regression and classification performance. The choice of evaluation metric often depends on the model's output format and the specific goals of the study—whether continuous value prediction or binary classification. As a result, direct comparisons between studies can be challenging, particularly when different endpoints or performance criteria are reported. This variability underscores the need for standardized evaluation frameworks in future work to better assess and compare model effectiveness.

B. Drawbacks and Limitations

Tomori et al. [49], Lam et al. [50], Ono et al. [51], Huang et al. [52], Wang et al. [53], and Song et al. [54] used the parameter prediction approach. Using a prediction approach, all studies indicated that machine learning models could be effectively trained using log files to predict machine parameters at the time of treatment delivery for new treatment plans. These studies vary in the models explored, including SVM, RF, CNNs, and others. All models have relatively promising accuracy as seen in Table I. However, Tomori et al.'s scope was limited to prostate IMRT plans, Huang et al. was limited to chest IMRT plans, and Song et al. was similarly limited to nasopharyngeal carcinoma and only used static gantry IMRT plans. Lam et al. included plans for multiple anatomical sites but were still specific to IMRT. Ono et al. and Wang et al. were specific to VMAT plans. Ono et al. and Lam et al. both performed their studies on multiple linear accelerators, but only Lam et al. used data from more than one institution. All six studies acknowledge that by using trajectory files, which are dependent on the linear accelerator itself, there is some vulnerability to machine-based error. As such, most log file-based PSQA is considered an enhancement to other QA measures that ensure the machine is calibrated appropriately, either with separate protocols or by incorporating additional sources of data into future models.

Error detection studies such as those by Kimura et al. [55], Sakai et al. [56], and Nyflot et al. [57] were similarly limited to one treatment plan type from a single institution. The only study that incorporated both VMAT and IMRT plans into a single study was an error detection study by Chuang et al. However, the study was only focused on MLC errors.

C. Positive Developments

These preliminary studies have gleaned significant insights into creating a holistic model for automating PSQA using log file data with a clear improvement upon methods over time. Lam et al. trained their model for predicting dosimetric effects in lieu of GPR to overcome any discrepancies between gamma index and errors that are clinically relevant [50]. Kimura et al. directly compared gamma map-based CNN models with dose difference map-based CNN models and found dose difference maps were more accurate [55]. Sakai et al. included radiomic data which resulted in higher sensitivity and specificity for MLC position and MLC modeling errors [56]. Hirashima et al. utilized a combination of 3D dosiomic features and plan complexity in a tree-based model [58]. Tomori et al.'s GPR prediction-based CNN model struggled with overestimating low GPR values and underestimating GPR in the test set [49][59]. Song et al. developed a novel model that weighed the MSE loss function to mitigate this class imbalance with promising results [54]. However, as all these studies have been limited to relatively small, single, or double institution datasets, their results are difficult to directly compare to one another. Additionally, most of the literature has been performed using Varian machines [27]. Although Varian machines are widely used in the US, Elekta machines are also used.

IV. Discussion

Literature has broadly indicated that CNNs and other Deep Learning models appear to be the most successful at creating a model that is robust against certain biases seen in SVM and tree-based algorithms [60]. Although some studies have utilized data augmentation, most studies have agreed that to bring these findings to a clinically relevant standpoint, sufficient data must be collected from multiple institutions, techniques, treatment machines, and anatomical sites [61][62]. Additionally, encompassing both Varian and Elekta machines is essential to ensure this PSQA strategy is accurate on both platforms [63].

Furthermore, past work has predominantly focused on deterministic methods, which are ideal for providing direct, quantitative evaluations of dose delivery accuracy. While these are incredibly important in the overall application of the model, there are many aspects of treatment that carry uncertainty. Error tolerance, dose assessments, and multi-criteria evaluations are all subject to imprecision. Cilla et al. approached these aspects by using a "traffic light" protocol [64]. The protocol leveraged plan complexity to designate plans as acceptable (green light), requires further verifcation (orange light), or unacceptable (red light) [64]. Fuzzy logic follows similar reasoning and has been successfully applied to radiation control systems and treatment plan optimization [65][66]. Fuzzy logic uses fuzzy sets and linguistic variables to model uncertain or imprecise information. Desired variables can be assigned degrees of truth rather than a yes/no value. When applied to complex systems, this mathematical system eliminates the restriction of binary values to create more human-like decision making. The Fuzzy-CID3 (F-CID3) algorithm is a tree-based, hybrid method that combines neural networks and fuzzy sets, generating its own topology. Using a neural fuzzy number tree with a class separation method, the F-CID3 algorithm simplifies architecture compared to precesssors, achieving better performance with fewer connections [67].

While fuzzy logic offers a way to model uncertainty in PSQA, it also highlights a broader need: the development of systems that can continuously adapt to changing clinical conditions. Future models must be not only accurate, but also adaptable, scalable, and self-evolving. Instead of relying on retraining static models each time conditions change, future systems should be capable of continuous learning and modular updates.

Incremental learning is one potential strategy where models are can be updated gradually as new data is introduced [68]. This avoids the need for complete retraining by using an adaptable models that can change in real-time. It is particularly well suited for large datasets that requires stability. However, these approaches are vulnerable to catastrophic forgetting where older knowledge can be lost when incorporating new data [69]. Techniques such as elastic weight consolidation can help address this issue by preserving important parameters, though they introduce new challenges in implementation and tuning. Other mitigation techniques include replay, template-based classification, and context dependent processing[70]–[72].

Fine-tuning offers another solution with further training on related datasets [73]. This could be particularly effective when adapting models to new machines, clinics, or delivery methods. It is less computationally demanding than full retraining, but fine-tuning must be handled carefully to avoid overfitting, especially in settings with limited or imbalanced data which are already inherent issues with log-file based PSQA models.

Genetic algorithms are another potential method for model evolution [74][75]. By using population-based search strategies, these algorithms can explore different architectures, hyperparameters, and feature selections over time. Genetic algorithms are computationally intensive which can make them difficult to implement. However, they are well suited for continuous optimization in non-critical processing environments and are not as prone to overfitting.

Figure 1 depicts a potential workflow that incorproates suggested improvments to prior models. These To address the limitations of prior log file-based IMRT PSQA modeling strategies, a revised workflow is proposed, as illustrated in Figure 1 Previous machine learning and deep learning models often relied on data from a single institution, which greatly limited model generalizability and increased the risk of overfitting due to site-specific bias. In the proposed framework, the input data is expanded to incorporate contributions from multiple institutions, enhancing the diversity of the training dataset and improving the model's capacity for generalizability. Additionally, the integration of a genetic algorithm with fuzzy logic is expected to improve feature selection and model optimization. The genetic algorithm enables a more robust exploration of potential feature sets, while fuzzy logic facilitates flexible decision-making in response to variability both within and across institutions. Together, these modifications are intended to yield models with not only increased predictive accuracy but also significantly enhanced clinical applicability by allowing the system to evolve alongside future clinical shifts.

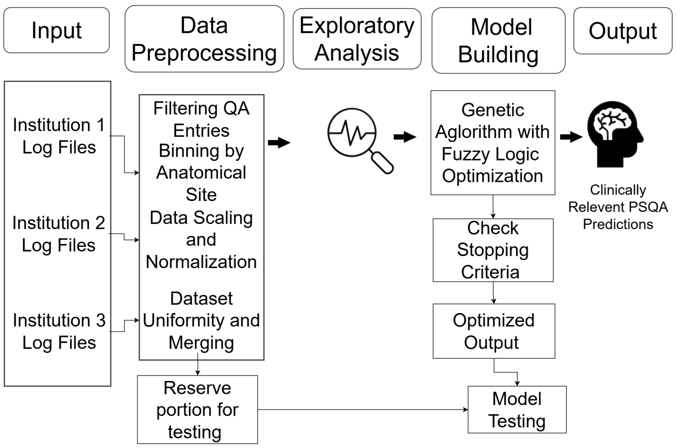


Figure 1. Suggested workflow for log file-based IMRT PSQA modeling. Enhancements include the incorporation of multi-institutional input data to improve model generalizability. The integration of a genetic algorithm with fuzzy logic is expected to enhance optimization, feature selection, and adaptability.

As a supplemental technique to the aforementioned models, federated learning can provide a network-preserving framework for scaling PSQA across multiple clinics [76]. Rather than sharing patient data, each institution trains a local model on its own dataset and transmits only model updates (e.g., gradients or weights) to a central server. The server then aggregates the updates into a global model and redistributes it. While federated learning effectively addresses data-sharing restrictions, its implementation presents challenges such as non-identically distributed data across sites, communication inefficiencies, and synchronization issues. Nevertheless, it remains a promising direction for building generalizable, robust QA models at scale.

V. Conclusions and Future Directions

Log file–based PSQA has emerged as a viable and efficient alternative to traditional phantom-based methods, particularly for IMRT and VMAT treatments. By leveraging machine learning models such as SVMs, tree-based algorithms, and CNNs, recent studies have demonstrated the ability to predict GPR outcomes and detect delivery errors directly from log file data. These methods offer a promising path toward scalable, real-time QA workflows that reduce clinical burden while maintaining—or even enhancing—treatment safety.

However, several challenges must be addressed before these tools can be widely implemented in clinical practice. Most current models rely on machine-reported parameters, limiting their ability to detect mechanical miscalibrations or TPS-related errors. In addition, the majority are trained on single-institution datasets, with limited anatomical and vendor diversity, which restricts their generalizability. These limitations highlight the need for multi-institutional datasets that reflect a broader spectrum of treatment techniques, patient populations, and machine types [54][64][77]–[79].

Future research must prioritize both generalizability and clinical adaptability. Multi-center collaborations that incorporate diverse planning protocols and hardware systems will be critical to developing robust, transferable models. Models must also be designed to remain effective in evolving clinical environments. Techniques such as incremental learning, transfer learning, and modular architectures can enable continuous model improvement without requiring full retraining. Federated learning also offers a promising privacy-preserving strategy for distributed model development across institutions.

Equally important, especially in the clinical setting, is the need for transparency and interpretability. Integrating explainable machine learning and deep learning tools can help clinicians understand how models generate predictions and identify which features contribute to error detection. This not only fosters trust and accountability, but also facilitates earlier, more targeted interventions.

Altogether, these advancements represent a shift toward faster, more efficient, and responsive QA systems. Future PSQA workflows should ideally evolve in step with technological innovation, while enhancing precision and safety in modern radiation oncology.

References

[1] K. De Jesus, L. Dunn, D. Thomas, and L. Sztandera, “Recent advances in machine learning for log file-based psqa for imrt and vmat”, in *Proceedings of the 2025 Second International Conference on AI-Health*, Mar. 2025, ISBN: 978-1-68558-247-0.

[2] R. L. Siegel, A. N. Giaquinto, and A. Jemal, “Cancer statistics, 2024”, *CA: A Cancer Journal for Clinicians*, vol. 74, no. 1, pp. 12–49, Jan. 2024. DOI: 10.3322/caac.21820. Accessed: Nov. 10, 2024. [Online]. Available: http://dx.doi.org/10.3322/caac.21820.

[3] A. A. for Cancer Research, *Aacr Cancer Progress Report 2024*. AACR, 2024, ISBN: 979-8-9857852-7-2. Accessed: Nov. 10, 2024.

[4] R. J. Hutten et al., “Worsening racial disparities in utilization of intensity modulated radiation therapy”, *Advances in Radiation Oncology*, vol. 7, no. 3, p. 100 887, Jan. 2022, ISSN: 24521094. DOI: 10.1016/j.adro.2021.100887. Accessed: Nov. 10, 2024. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2452109421002451.

[5] A. Agnew, C. E. Agnew, M. W. D. Grattan, A. R. Hounsell, and C. K. McGarry, “Monitoring daily MLC positional errors using trajectory log files and EPID measurements for IMRT and VMAT deliveries”, *Physics in Medicine and Biology*, vol. 59, no. 9, N49–63, May 2014. DOI: 10.1088/0031-9155/59/9/N49. Accessed: Oct. 10, 2024. [Online]. Available: http://dx.doi.org/10.1088/0031-9155/59/9/N49.

[6] C. E. Agnew, D. M. Irvine, and C. K. McGarry, “Correlation of phantom-based and log file patient-specific QA with complexity scores for VMAT”, *Journal of Applied Clinical Medical Physics / American College of Medical Physics*, vol. 15, no. 6, p. 4994, Sep. 2014. DOI: 10.1120/jacmp.v15i6.4994. Accessed: Oct. 24, 2024. [Online]. Available: http://dx.doi.org/10.1120/jacmp.v15i6.4994.

[7] N. Childress, Q. Chen, and Y. Rong, "Parallel/opposed: IMRT QA using treatment log files is superior to conventional measurement-based method", *Journal of Applied Clinical Medical Physics / American College of Medical Physics*, vol. 16, no. 1, p. 5385, Jan. 2015. DOI: 10.1120/jacmp.v16i1.5385. Accessed: Oct. 10, 2024. [Online]. Available: http://dx.doi.org/10.1120/jacmp.v16i1.5385.

[8] B. Neal et al., "A clinically observed discrepancy between image-based and log-based MLC positions", *Medical Physics*, vol. 43, no. 6, pp. 2933–2935, Jun. 2016. DOI: 10.1118/1.4949002. Accessed: Oct. 24, 2024. [Online]. Available: http://dx.doi.org/10.1118/1.4949002.

[9] M. Barnes et al., "Insensitivity of machine log files to MLC leaf backlash and effect of MLC backlash on clinical dynamic MLC motion: An experimental investigation", *Journal of Applied Clinical Medical Physics / American College of Medical Physics*, vol. 23, no. 9, e13660, Sep. 2022. DOI: 10.1002/acm2.13660. Accessed: Oct. 24, 2024. [Online]. Available: http://dx.doi.org/10.1002/acm2.13660.

[10] J. Waring, C. Lindvall, and R. Umeton, "Automated machine learning: Review of the state-of-the-art and opportunities for healthcare", *Artificial Intelligence in Medicine*, vol. 104, p. 101 822, Apr. 2020, ISSN: 09333657. DOI: 10.1016/j.artmed.2020.101822. Accessed: Jan. 22, 2025. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0933365719310437.

[11] A. Esteva et al., "A guide to deep learning in healthcare", *Nature Medicine*, vol. 25, no. 1, pp. 24–29, Jan. 2019, ISSN: 1078-8956. DOI: 10.1038/s41591-018-0316-z. Accessed: Oct. 20, 2021. [Online]. Available: http://www.nature.com/articles/s41591-018-0316-z.

[12] K. Rasheed et al., "Explainable, trustworthy, and ethical machine learning for healthcare: A survey", *Computers in Biology and Medicine*, vol. 149, p. 106 043, Oct. 2022. DOI: 10.1016/j.compbiomed.2022.106043. Accessed: Jan. 22, 2025. [Online]. Available: http://dx.doi.org/10.1016/j.compbiomed.2022.106043.

[13] K. Y. Ngiam and I. W. Khor, "Big data and machine learning algorithms for health-care delivery", *The Lancet Oncology*, vol. 20, no. 5, e262–e273, May 2019. DOI: 10.1016/S1470-2045(19)30149-4. Accessed: Jan. 22, 2025. [Online]. Available: http://dx.doi.org/10.1016/S1470-2045(19)30149-4.

[14] J. M. Park, J.-I. Kim, and H.-G. Wu, "Technological advances in charged-particle therapy", *Cancer Research and Treatment : Official Journal of Korean Cancer Association*, vol. 53, no. 3, pp. 635–640, Jul. 2021. DOI: 10.4143/crt.2021.706. Accessed: Jan. 22, 2025. [Online]. Available: http://dx.doi.org/10.4143/crt.2021.706.

[15] R. A. Chandra, F. K. Keane, F. E. M. Voncken, and C. R. Thomas, "Contemporary radiotherapy: Present and future", *The Lancet*, vol. 398, no. 10295, pp. 171–184, Jul. 2021. DOI: 10.1016/S0140-6736(21)00233-6. Accessed: Jan. 22, 2025. [Online]. Available: http://dx.doi.org/10.1016/S0140-6736(21)00233-6.

[16] J. Bertholet, Y. Vinogradskiy, Y. Hu, and D. J. Carlson, "Advances in image-guided adaptive radiation therapy", *International Journal of Radiation Oncology, Biology, Physics*, vol. 110, no. 3, pp. 625–628, Jul. 2021. DOI: 10.1016/j.ijrobp.2021.02.047. Accessed: Jan. 22, 2025. [Online]. Available: http://dx.doi.org/10.1016/j.ijrobp.2021.02.047.

[17] S. Bartzsch et al., "Technical advances in x-ray microbeam radiation therapy", *Physics in Medicine and Biology*, vol. 65, no. 2, 02TR01, Jan. 2020. DOI: 10.1088/1361-6560/ab5507. Accessed: Jan. 22, 2025. [Online]. Available: http://dx.doi.org/10.1088/1361-6560/ab5507.

[18] C. X. Yu, C. J. Amies, and M. Svatos, "Planning and delivery of intensity-modulated radiation therapy", *Medical Physics*, vol. 35, no. 12, pp. 5233–5241, Sep. 2008. DOI: 10.1118/1.3002305. Accessed: Oct. 10, 2024. [Online]. Available: http://dx.doi.org/10.1118/1.3002305.

[19] A. F. I. Osman, N. M. Maalej, and K. Jayesh, "Prediction of the individual multileaf collimator positional deviations during dynamic IMRT delivery priori with artificial neural network", *Medical Physics*, vol. 47, no. 4, pp. 1421–1430, Apr. 2020. DOI: 10.1002/mp.14014. Accessed: Oct. 20, 2024. [Online]. Available: http://dx.doi.org/10.1002/mp.14014.

[20] P. Szeverinski et al., "Error sensitivity of a log file analysis tool compared with a helical diode array dosimeter for VMAT delivery quality assurance", *Journal of Applied Clinical Medical Physics / American College of Medical Physics*, vol. 21, no. 11, pp. 163–171, Sep. 2020. DOI: 10.1002/acm2.13051. Accessed: Oct. 10, 2024. [Online]. Available: http://dx.doi.org/10.1002/acm2.13051.

[21] E. E. Klein et al., "Task group 142 report: Quality assurance of medical accelerators", *Medical Physics*, vol. 36, no. 9, pp. 4197–4212, Sep. 2009. DOI: 10.1118/1.3190392. Accessed: Jul. 1, 2020. [Online]. Available: http://dx.doi.org/10.1118/1.3190392.

[22] D. A. Low, W. B. Harms, S. Mutic, and J. A. Purdy, "A technique for the quantitative evaluation of dose distributions", *Medical Physics*, vol. 25, no. 5, pp. 656–661, May 1998. DOI: 10.1118/1.598248.

[23] D. A. Low and J. F. Dempsey, "Evaluation of the gamma dose distribution comparison method", *Medical Physics*, vol. 30, no. 9, pp. 2455–2464, Sep. 2003. DOI: 10.1118/1.1598711.

[24] B. E. Nelms, H. Zhen, and W. A. Tomé, "Per-beam, planar imrt qa passing rates do not predict clinically relevant patient dose errors", *Medical Physics*, vol. 38, no. 2, pp. 1037–1044, Feb. 2011. DOI: 10.1118/1.3544657.

[25] B. E. Nelms et al., "Evaluating imrt and vmat dose accuracy: Practical examples of failure to detect systematic errors when applying a commonly used metric and action levels", *Medical Physics*, vol. 40, no. 11, p. 111 722, Sep. 2013. DOI: 10.1118/1.4826166.

[26] T. Depuydt, A. Van Esch, and D. P. Huyskens, "A quantitative evaluation of imrt dose distributions: Refinement and clinical assessment of the gamma evaluation", *Radiotherapy and Oncology*, vol. 62, no. 3, pp. 309–319, Mar. 2002. DOI: 10.1016/s0167-8140(01)00497-2.

[27] G. A. Ezzell et al., "IMRT commissioning: Multiple institution planning and dosimetry comparisons, a report from AAPM task group 119", *Medical Physics*, vol. 36, no. 11, pp. 5359–5373, Sep. 2009. DOI: 10.1118/1.3238104. Accessed: Oct. 20, 2024. [Online]. Available: http://dx.doi.org/10.1118/1.3238104.

[28] M. Miften et al., "Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM task group no. 218", *Medical Physics*, vol. 45, no. 4, e53–e83, Apr. 2018. DOI: 10.1002/mp.12810. Accessed: Oct. 20, 2024. [Online]. Available: http://dx.doi.org/10.1002/mp.12810.

[29] R. A. C. Siochi, A. Molineu, and C. G. Orton, "Point/counterpoint. patient-specific QA for IMRT should be performed using software rather than hardware methods", *Medical Physics*, vol. 40, no. 7, p. 070 601, Jul. 2013. DOI: 10.1118/1.4794929. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1118/1.4794929.

[30] J. C. Smith, S. Dieterich, and C. G. Orton, "Point/counterpoint. it is still necessary to validate each individual IMRT treatment plan with dosimetric measurements before delivery", *Medical Physics*, vol. 38, no. 2, pp. 553–555, Feb. 2011. DOI: 10.1118/1.3512801. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1118/1.3512801.

[31] D. W. Litzenberg, J. M. Moran, and B. A. Fraass, "Verification of dynamic and segmental IMRT delivery by dynamic log file analysis", *Journal of Applied Clinical Medical Physics /*

American College of Medical Physics, vol. 3, no. 2, pp. 63–72, 2002. DOI: 10.1120/jacmp.v3i2.2578. Accessed: Jan. 22, 2025. [Online]. Available: http://dx.doi.org/10.1120/jacmp.v3i2.2578.

[32] N. L. Childress, C. Bloch, R. A. White, M. Salehpour, and I. I. Rosen, "Detection of imrt delivery errors using a quantitative 2d dosimetric verification system", *Medical Physics*, vol. 32, no. 1, pp. 153–162, Jan. 2005. DOI: 10.1118/1.1829171.

[33] J. M. Moran et al., "Safety considerations for IMRT: Executive summary", *Practical Radiation Oncology*, vol. 1, no. 3, pp. 190–195, Sep. 2011. DOI: 10.1016/j.prro.2011.04.008. Accessed: Oct. 10, 2024. [Online]. Available: http://dx.doi.org/10.1016/j.prro.2011.04.008.

[34] A. L. McNiven, M. B. Sharpe, and T. G. Purdie, "A new metric for assessing IMRT modulation complexity and plan deliverability", *Medical Physics*, vol. 37, no. 2, pp. 505–515, Feb. 2010. DOI: 10.1118/1.3276775. Accessed: Oct. 10, 2024. [Online]. Available: http://dx.doi.org/10.1118/1.3276775.

[35] L. Masi, R. Doro, V. Favuzza, S. Cipressi, and L. Livi, "Impact of plan parameters on the dosimetric accuracy of volumetric modulated arc therapy", *Medical Physics*, vol. 40, no. 7, p. 071 718, Jul. 2013. DOI: 10.1118/1.4810969. Accessed: Oct. 10, 2024. [Online]. Available: http://dx.doi.org/10.1118/1.4810969.

[36] B. Sun et al., "Initial experience with TrueBeam trajectory log files for radiation therapy delivery verification", *Practical Rdiation Oncology*, vol. 3, no. 4, e199–208, Dec. 2013. DOI: 10.1016/j.prro.2012.11.013. Accessed: Jan. 22, 2025. [Online]. Available: http://dx.doi.org/10.1016/j.prro.2012.11.013.

[37] C. Cortes and V. Vapnik, "Support-vector networks", *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995. DOI: 10.1007/BF00994018.

[38] T. K. Ho, "Random decision forests", in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, vol. 1, 1995, 278–282 vol.1. DOI: 10.1109/ICDAR.1995.598994.

[39] J. H. Friedman, "Greedy function approximation: A gradient boosting machine", *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001, ISSN: 0090-5364. DOI: 10.1214/aos/1013203451. Accessed: Jul. 5, 2020. [Online]. Available: http://projecteuclid.org/euclid.aos/1013203451.

[40] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system", in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, New York, New York, USA: ACM Press, Aug. 2016, pp. 785–794, ISBN: 9781450342322. DOI: 10.1145/2939672.2939785. Accessed: Jan. 22, 2025. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2939672.2939785.

[41] R. Tibshirani, "Regression shrinkage and selection via the lasso", *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, Jan. 1996, ISSN: 00359246. DOI: 10.1111/j.2517-6161.1996.tb02080.x. Accessed: Jun. 10, 2016. [Online]. Available: http://doi.wiley.com/10.1111/j.2517-6161.1996.tb02080.x.

[42] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems", *Technometrics : A Journal of Statistics for the Physical, Chemical, and Engineering Sciences*, vol. 12, no. 1, pp. 55–67, Feb. 1970, ISSN: 0040-1706. DOI: 10.1080/00401706.1970.10488634. Accessed: Jan. 22, 2025. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/00401706.1970.10488634.

[43] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity", *The Bulletin of Mathematical Biophysics*, vol. 5, pp. 115–133, 1943. DOI: 10.1007/BF02478259.

[44] K. Fukushima, "Visual feature extraction by a multilayered network of analog threshold elements", *IEEE Transactions on Systems Science and Cybernetics*, vol. 5, no. 4, pp. 322–333, 1969. DOI: 10.1109/TSSC.1969.300225.

[45] A. G. Ivakhnenko, "The group method of data handling—a rival of the method of stochastic approximation", *Soviet Automatic Control*, vol. 1, no. 3, pp. 43–55, 1968.

[46] J. Schmidhuber, "Deep learning in neural networks: An overview", *Neural Networks*, vol. 61, pp. 85–117, Jan. 2015. DOI: 10.1016/j.neunet.2014.09.003.

[47] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition", *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. DOI: 10.1109/5.726791.

[48] W. Ding, M. Abdel-Basset, H. Hawash, and A. M. Ali, "Explainability of artificial intelligence methods, applications and challenges: A comprehensive survey", *Information Sciences*, vol. 615, pp. 238–292, 2022, ISSN: 0020-0255. DOI: https://doi.org/10.1016/j.ins.2022.10.013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S002002552201132X.

[49] S. Tomori et al., "A deep learning-based prediction model for gamma evaluation in patient-specific quality assurance", *Medical Physics*, Jul. 2018. DOI: 10.1002/mp.13112. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1002/mp.13112.

[50] D. Lam et al., "Predicting gamma passing rates for portal dosimetry-based IMRT QA using machine learning", *Medical Physics*, vol. 46, no. 10, pp. 4666–4675, Oct. 2019. DOI: 10.1002/mp.13752. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1002/mp.13752.

[51] T. Ono et al., "Prediction of dosimetric accuracy for VMAT plans using plan complexity parameters via machine learning", *Medical Physics*, vol. 46, no. 9, pp. 3823–3832, Sep. 2019. DOI: 10.1002/mp.13669. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1002/mp.13669.

[52] Y. Huang et al., "Deep learning for patient-specific quality assurance: Predicting gamma passing rates for IMRT based on delivery fluence informed by log files", *Technology in Cancer Research & Treatment*, vol. 21, p. 15 330 338 221 104 881, 2022. DOI: 10.1177/15330338221104881. Accessed: Nov. 3, 2024. [Online]. Available: http://dx.doi.org/10.1177/15330338221104881.

[53] L. Wang et al., "Multi-task autoencoder based classification-regression model for patient-specific VMAT QA", *Physics in Medicine and Biology*, vol. 65, no. 23, p. 235 023, Sep. 2020. DOI: 10.1088/1361-6560/abb31c. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1088/1361-6560/abb31c.

[54] W. Song et al., "Improving the performance of deep learning models in predicting and classifying gamma passing rates with discriminative features and a class balancing technique: A retrospective cohort study", *Radiation Oncology*, vol. 19, no. 1, p. 98, Jul. 2024. DOI: 10.1186/s13014-024-02496-5. Accessed: Nov. 3, 2024. [Online]. Available: http://dx.doi.org/10.1186/s13014-024-02496-5.

[55] Y. Kimura, N. Kadoya, S. Tomori, Y. Oku, and K. Jingu, "Error detection using a convolutional neural network with dose difference maps in patient-specific quality assurance for volumetric modulated arc therapy", *Physica Medica : PM : An International Journal Devoted to the Applications of Physics to Medicine and Biology : Official Journal of the Italian Association of Biomedical Physics (AIFB)*, vol. 73, pp. 57–64, May 2020. DOI: 10.1016/j.ejmp.2020.03.022. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1016/j.ejmp.2020.03.022.

[56] M. Sakai et al., "Detecting MLC modeling errors using radiomics-based machine learning in patient-specific QA with an EPID for intensity-modulated radiation therapy", *Medical Physics*, vol. 48, no. 3, pp. 991–1002, Mar. 2021. DOI: 10.

1002/mp.14699. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1002/mp.14699.

[57] M. J. Nyflot, P. Thammasorn, L. S. Wootton, E. C. Ford, and W. A. Chaovalitwongse, "Deep learning for patient-specific quality assurance: Identifying errors in radiotherapy delivery by radiomic analysis of gamma images with convolutional neural networks", *Medical Physics*, vol. 46, no. 2, pp. 456–464, Feb. 2019. DOI: 10.1002/mp.13338. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1002/mp.13338.

[58] H. Hirashima et al., "Improvement of prediction and classification performance for gamma passing rate by using plan complexity and dosiomics features", *Radiotherapy and Oncology*, vol. 153, pp. 250–257, Dec. 2020, ISSN: 01678140. DOI: 10.1016/j.radonc.2020.07.031. Accessed: Jan. 17, 2025. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167814020306769.

[59] S. Tomori et al., "Systematic method for a deep learning-based prediction model for gamma evaluation in patient-specific quality assurance of volumetric modulated arc therapy", *Medical Physics*, vol. 48, no. 3, pp. 1003–1018, Mar. 2021. DOI: 10.1002/mp.14682. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1002/mp.14682.

[60] A. F. I. Osman and N. M. Maalej, "Applications of machine and deep learning to patient-specific IMRT/VMAT quality assurance", *Journal of Applied Clinical Medical Physics / American College of Medical Physics*, vol. 22, no. 9, pp. 20–36, Sep. 2021. DOI: 10.1002/acm2.13375. Accessed: Sep. 23, 2024. [Online]. Available: http://dx.doi.org/10.1002/acm2.13375.

[61] Y. Interian et al., "Deep nets vs expert designed features in medical physics: An IMRT QA case study", *Medical Physics*, vol. 45, no. 6, pp. 2672–2680, Jun. 2018. DOI: 10.1002/mp.12890. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1002/mp.12890.

[62] D. M. Fondevila, P. J. Rios, D. M. D. Peñalva, and S. Arbiser, "Predicting gamma passing rates for portal dosimetry-based IMRT QA using deep learning", *International Journal of Radiation Oncology, Biology, Physics*, vol. 111, no. 3S, e110–e111, Sep. 2021. DOI: 10.1016/j.ijrobp.2021.07.515. Accessed: Jan. 22, 2025. [Online]. Available: http://dx.doi.org/10.1016/j.ijrobp.2021.07.515.

[63] I. J. Das et al., "Accelerator beam data commissioning equipment and procedures: Report of the TG-106 of the therapy physics committee of the AAPM", *Medical Physics*, vol. 35, no. 9, pp. 4186–4215, Sep. 2008. DOI: 10.1118/1.2969070. Accessed: Jan. 22, 2025. [Online]. Available: http://dx.doi.org/10.1118/1.2969070.

[64] S. Cilla et al., "Prediction and classification of VMAT dosimetric accuracy using plan complexity and log-files analysis", *Physica Medica : PM : an international journal devoted to the applications of physics to medicine and biology : Official Journal of the Italian Association of Biomedical Physics (AIFB)*, vol. 103, pp. 76–88, Sep. 2022. DOI: 10.1016/j.ejmp.2022.10.004. Accessed: Nov. 3, 2024. [Online]. Available: http://dx.doi.org/10.1016/j.ejmp.2022.10.004.

[65] T.-F. Lee et al., "A fuzzy system for evaluating radiation treatment plans of head and neck cancer", in *2012 9th International Conference on Fuzzy Systems and Knowledge Discovery*, 2012, pp. 510–514. DOI: 10.1109/FSKD.2012.6234043.

[66] C. Pinter, T. Olding, L. J. Schreiner, and G. Fichtinger, "Using fuzzy logics to determine optimal oversampling factor for voxelizing 3D surfaces in radiation therapy", *Soft Computing*, vol. 24, no. 24, pp. 18 959–18 970, Dec. 2020, ISSN: 1432-7643. DOI: 10.1007/s00500-020-05126-w. Accessed: Jan. 22, 2025. [Online]. Available: https://link.springer.com/10.1007/s00500-020-05126-w.

[67] L. Sztandera, "Computational intelligence foundations", in *Computational Intelligence in Business Analytics: Concepts, Methods, and Tools for Big Data Applications*, 1st ed., Upper Saddle River, New Jersey: Pearson, 2014, ISBN: 0-13-355208-X. Accessed: Jan. 18, 2025.

[68] G. M. van de Ven and A. S. Tolias, "Three types of incremental learning", *Nature Machine Intelligence*, vol. 4, no. 12, pp. 1185–1197, Dec. 2022. DOI: 10.1038/s42256-022-00568-3.

[69] J. Kirkpatrick et al., "Overcoming catastrophic forgetting in neural networks", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, no. 13, pp. 3521–3526, Mar. 2017. DOI: 10.1073/pnas.1611835114.

[70] H. Shin, J. K. Lee, J. Kim, and J. Kim, "Continual learning with deep generative replay", in *Advances in Neural Information Processing Systems*, I. Guyon et al., Eds., vol. 30, Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/0efbe98067c6c73dba1250d2beaa81f9-Paper.pdf.

[71] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, "Icarl: Incremental classifier and representation learning", in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5533–5542. DOI: 10.1109/CVPR.2017.587.

[72] N. Y. Masse, G. D. Grant, and D. J. Freedman, "Alleviating catastrophic forgetting using context-dependent gating and synaptic stabilization", *Proceedings of the National Academy of Sciences*, vol. 115, no. 44, E10467–E10475, 2018. DOI: 10.1073/pnas.1803839115. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1803839115. [Online]. Available: https://www.pnas.org/doi/abs/10.1073/pnas.1803839115.

[73] S. J. Pan and Q. Yang, "A survey on transfer learning", *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010. DOI: 10.1109/TKDE.2009.191.

[74] "Genetic algorithms and machine learning", *Machine Learning*, vol. 3, no. 2–3, pp. 95–99, Oct. 1988. DOI: 10.1007/BF00113892.

[75] S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: Past, present, and future", *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 8091–8126, 2021. DOI: 10.1007/s11042-020-10139-6.

[76] P. Kairouz et al., "Advances and open problems in federated learning", *Found. Trends Mach. Learn.*, vol. 14, no. 1–2, pp. 1–210, Jun. 2021, ISSN: 1935-8237. DOI: 10.1561/2200000083. [Online]. Available: https://doi.org/10.1561/2200000083.

[77] P. Viola et al., "Prediction of VMAT delivery accuracy using plan modulation complexity score and log-files analysis", *Biomedical Physics & Engineering Express*, vol. 8, no. 5, Aug. 2022. DOI: 10.1088/2057-1976/ac82c6. Accessed: Nov. 3, 2024. [Online]. Available: http://dx.doi.org/10.1088/2057-1976/ac82c6.

[78] Y. Huang et al., "A feasibility study to predict 3D dose delivery accuracy for IMRT using DenseNet with log files", *Journal of X-ray Science and Technology*, vol. 32, no. 4, pp. 1199–1208, 2024. DOI: 10.3233/{XST}-230412. Accessed: Nov. 3, 2024. [Online]. Available: http://dx.doi.org/10.3233/%7BXST%7D-230412.

[79] T. Ono et al., "Applications of artificial intelligence for machine- and patient-specific quality assurance in radiation therapy: Current status and future directions", *Journal of Radiation Research*, vol. 65, no. 4, pp. 421–432, Jul. 2024. DOI: 10.1093/jrr/rrae033. Accessed: Sep. 23, 2024. [Online]. Available: http://dx.doi.org/10.1093/jrr/rrae033.

[80] J. N. K. Carlson et al., "A machine learning approach to the accurate prediction of multi-leaf collimator positional errors", *Physics in Medicine and Biology*, vol. 61, no. 6, pp. 2514–2531, Mar. 2016. DOI: 10.1088/0031-9155/61/6/2514. Accessed:

Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1088/0031-9155/61/6/2514.

[81] D. A. Granville, J. G. Sutherland, J. G. Belec, and D. J. La Russa, “Predicting VMAT patient-specific QA results using a support vector classifier trained on treatment plan characteristics and linac QC metrics”, *Physics in Medicine and Biology*, vol. 64, no. 9, p. 095 017, Apr. 2019. DOI: 10.1088/1361-6560/ab142e. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1088/1361-6560/ab142e.

[82] C. Ma et al., “The structural similarity index for IMRT quality assurance: Radiomics-based error classification”, *Medical Physics*, vol. 48, no. 1, pp. 80–93, Jan. 2021. DOI: 10.1002/mp.14559. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1002/mp.14559.

[83] P. D. Wall and J. D. Fontenot, “Application and comparison of machine learning models for predicting quality assurance outcomes in radiation therapy treatment planning”, *Informatics in Medicine Unlocked*, vol. 18, p. 100 292, 2020, ISSN: 23529148. DOI: 10.1016/j.imu.2020.100292. Accessed: Jan. 17, 2025. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2352914819303661.

[84] K.-C. Chuang, W. Giles, and J. Adamson, “A tool for patient-specific prediction of delivery discrepancies in machine parameters using trajectory log files”, *Medical Physics*, vol. 48, no. 3, pp. 978–990, Mar. 2021. DOI: 10.1002/mp.14670. Accessed: Oct. 20, 2024. [Online]. Available: http://dx.doi.org/10.1002/mp.14670.

[85] K. S. Lew et al., “Prediction of portal dosimetry quality assurance results using log files-derived errors and machine learning techniques”, *Frontiers in oncology*, vol. 12, p. 1 096 838, 2022. DOI: 10.3389/fonc.2022.1096838. Accessed: Nov. 3, 2024. [Online]. Available: http://dx.doi.org/10.3389/fonc.2022.1096838.

TABLE I. SUMMARY OF RECENT STUDIES USING MACHINE LEARNING MODELS FOR IMRT/VMAT PSQA. (AUC= AREA UNDER THE CURVE, MAE= MEAN ABSOLUTE ERROR, RMSE= ROOT MEAN SQUARE ERROR, SR= SPEARMAN'S RANK CORRELATION COEFFICIENT)

Author/Year	Plan Type	Dataset Size	Anatomic Sites	Algorithm	QA Outcome	Feature Count	Key Results
Carlson et al. 2016 [80]	VMAT	74 plans (3,161,280 data points)	Multiple	RF	Error detection	6	RMSE= 0.193mm (linear regression)
Tomori et al. 2018 [49]	IMRT	60 plans	Prostate	CNN	Parameter prediction	N/A	Errors within 1.10% at 3%/3mm criteria
Interian et al. 2018 [61]	IMRT	498 plans	Multiple	CNN	Parameter prediction	N/A	MAE= 0.70% at 3%/3mm criteria
Lam et al. 2019 [50]	IMRT	1497 beams	Multiple	Tree-based	Parameter prediction	31	Errors within 3% for 98% of predictions at 2%/2mm criteria
Ono et al. 2019 [51]	VMAT	600 plans	Multiple	Regression Tree, ANN, Other	Parameter prediction	28	Mean prediction error= -0.2% at 3%/3mm criteria (ANN)
Granville et al. 2019 [81]	VMAT	1,620 beams	Multiple	SVM	Error detection	60	AUC=0.88 (macro-averaged)
Nyflot et al. 2019 [57]	IMRT	186 beams (558 images)	Multiple	SVM, Decision Tree, Other	Error detection	145	Accuracy= 64.3% for SVM
Ma et al. 2020 [82]	IMRT	180 beams (1,620 images)	Multiple	SVM, RF, Other	Error detection	276	AUC=0.86 for linear SVM
Osman et al. 2020 [19]	IMRT	10 plans (360,800 datapoints)	Multiple	ANN	Error detection	14	RMSE=0.0096mm
Wall and Fontenot 2020 [83]	VMAT	500 plans	Multiple	SVM, Tree-Based, ANN	Parameter prediction	241	MAE=3.75% at 3%/3mm criteria (SVM)
Hirashima et al. 2020 [58]	VMAT	1,255 plans	Multiple	Tree-based	Parameter prediction	875	MAE=4.2% and AUC=0.83 at 2%/2mm criteria
Wang et al. 2020 [53]	VMAT	276 Plans	Multiple	ANN	Parameter prediction	N/A	Absolute prediction error=1.76% at 3%/3mm criteria
Kimura et al. 2020 [55]	VMAT	161 Beams	Prostate	CNN	Error detection	54	Accuracy=0.94
Tomori et al. 2020 [59]	VMAT	147 plans	Multiple	CNN	Parameter prediction	N/A	MAE=0.63% at 3%/3mm criteria
Sakai et al. 2021 [56]	IMRT	38 beams (152 error plans)	Multiple	SVM, Tree-based, Other	Error detection	837	AUC=1.00 for leaf transmission factor error, 1.0 for dosimetric leaf gap error, 0.80 for leaf positional error vs. error free (SVM)
Chuang et al. 2021 [84]	IMRT/VMAT	267 IMRT and VMAT plans (10,584,120 data points)	Multiple	Tree-based, Other	Error detection	7	RMSE=0.0085 mm (Boosted Tree Model)
Huang et al. 2022 [52]	IMRT	112 plans	Chest	CNN	Parameter prediction	4	MAE and RMSE decreased with stricter gamma criteria, while SR and R^2 increased as gamma criteria were made stricter (3%/3mm, 3%/2mm, 2%/3mm, and 2%/2mm)
Cilla et al. 2022 [64]	VMAT	651 plans/1,302 arcs	Multiple	SVM, Other	Parameter prediction	3	Precision of 93.1 for gamma % and 92.7% for gamma mean for the testing dataset at 2%/2mm (SVM)
Lew et al. 2022 [85]	VMAT	578 log files	Multiple	RF, SVM, Other	Parameter prediction	13	Average error of less than 2% with 1%/1mm criteria.
Song et al. 2024 [54]	IMRT	204 plans/2,348 fields	Nasopharyngeal Carcinoma	CNN	Parameter prediction	1-8	AUC= 0.92 with 0.77 sensitivity and 0.89 specificity