RESEARCH PAPER

Explainable and Uncertainty-Aware Federated Multimodal Foundation Intelligence for Privacy-Preserving and Trustworthy Medical Diagnosis

*[1]Burra Ramanuja Srinivas, [2]Dr Bhavana Jamalpur, [3]Raghavendra Rao Ankam
[1]Reseach Scholar, School of CS & AI, SR University, Warangal, Telangana, India. ORCID:0009-0002-0896-6347
[2]Associate Professor, School of CS & AI, SR University, Warangal, Telangana,India. ORCID:0000-0001-8454-3384.
[3]Associate Professor, RV Institute of Technology, Chebrolu, Guntur, Andhra Pradesh, India. ORCID: 0000-0002-87046073

Abstract

Modern healthcare increasingly depends on information from different sources, including medical images, electronic health records, laboratory results, and clinical notes. Bringing these sources together can provide a more complete picture of a patient and help clinicians make better and earlier diagnostic decisions. However, the use of such sensitive information also raises important concerns regarding patient privacy, data sharing, reliability of predictions, and the ability of clinicians to understand the reasons behind a diagnostic result. This study proposes an **Explainable and Uncertainty-Aware Federated Multimodal Foundation Intelligence (EU-FMFI) framework** for supporting secure, reliable, and trustworthy medical diagnosis. The framework allows healthcare institutions to learn collaboratively from diverse medical data while keeping patient information within the institution where it was collected. Information from medical images, electronic health records, and clinical notes is combined to obtain a more comprehensive understanding of individual patient conditions. The framework also considers the confidence associated with each diagnostic prediction, helping to identify cases where the available information may be insufficient or the diagnosis may be uncertain. In addition, the system provides understandable explanations that indicate the important clinical information contributing to a particular diagnostic decision. This can help doctors review the results, compare them with clinical evidence, and make informed decisions rather than relying solely on an automated prediction. The proposed approach is designed to address several important requirements of modern healthcare, including patient privacy, effective use of diverse medical information, reliable predictions, understandable results, and robustness across different healthcare institutions. Its performance can be evaluated using diagnostic measures such as accuracy, precision, recall, F1-score, and AUROC, together with measures of prediction confidence, calibration, privacy protection, communication efficiency, and explanation quality. Overall, the framework aims to provide a practical foundation for developing secure, transparent, and dependable diagnostic support systems that can assist healthcare professionals in making informed clinical decisions while protecting sensitive patient information.

Keywords: Federated Learning; Multimodal Medical Data; Explainable Diagnosis; Uncertainty Estimation; Patient Privacy; Vision Transformers; Clinical Language Models; Trustworthy Healthcare.

How to cite this article: *1Burra Ramanuja Srinivas 2Dr Bhavana Jamalpur 3Raghavendra Rao Ankam. Explainable and Uncertainty-Aware Federated Multimodal Foundation Intelligence for Privacy-Preserving and Trustworthy Medical Diagnosis. Int J Drug Deliv Technol. 2026;16(77s):963-968. DOI: 10.25258/ijddt.16.77s.112

Introduction

1.1 Background and Motivation : Healthcare systems generate large amounts of information from different sources, including medical images, electronic health records (EHRs), laboratory investigations, clinical observations, pathology reports, and patient histories. Each source provides a different view of a patient's health condition. Medical images can reveal structural and pathological changes, while EHRs and clinical records provide information about symptoms, previous illnesses, medications, laboratory findings, and treatment history. When these sources are considered together, they can provide a more complete understanding of a patient's condition than relying on a single source alone. This is particularly important for complex diseases such as cancer, cardiovascular disorders, neurological diseases, and other conditions in which diagnosis depends on several clinical indicators. The increasing availability of digital healthcare information therefore creates an opportunity to develop diagnostic systems that can assist clinicians in identifying disease patterns more accurately and at an earlier stage. Despite these opportunities, the use of healthcare data for collaborative diagnosis presents

several practical and ethical challenges. Medical information is highly sensitive, and hospitals and healthcare organizations are generally required to protect patient confidentiality and comply with applicable privacy regulations. Directly transferring patient records or medical images between institutions may therefore be difficult or inappropriate. In addition, medical data collected at different hospitals can vary considerably because of differences in patient populations, imaging equipment, clinical practices, recording procedures, and data quality. A diagnostic system that performs well in one institution may consequently provide less reliable results in another. These challenges highlight the need for a framework that can make effective use of diverse medical information while reducing unnecessary movement of sensitive patient data and maintaining reliable diagnostic performance across healthcare environments.

1.2. Need for Multimodal and Collaborative Diagnosis

Traditional diagnostic approaches often depend on individual data sources or isolated clinical assessments. Although medical imaging plays an important role in detecting abnormalities, images alone may not provide sufficient information to distinguish between similar diseases or to understand the patient's complete clinical background. Similarly, structured clinical records may contain valuable information but may not adequately represent visual characteristics observed in radiology or pathology images. Combining information from multiple sources can overcome some of these limitations by allowing complementary evidence to be considered during diagnosis. For example, an image may indicate a suspicious lesion, while laboratory findings, patient history, and clinical observations can provide additional evidence about its possible nature. A multimodal approach can therefore support a broader and more clinically meaningful assessment of disease. Collaborative learning across healthcare institutions offers another important opportunity. Individual hospitals may have limited numbers of patients or may possess only particular types of medical data. By learning from information available across multiple institutions without requiring the routine exchange of raw patient records, healthcare organizations can potentially develop more representative diagnostic models. Such collaboration is especially valuable for rare diseases and conditions in which individual institutions may not have sufficient cases for comprehensive analysis. However, collaborative approaches must carefully address differences among institutions, communication requirements, privacy risks, and variations in data quality. The proposed **Explainable and Uncertainty-Aware Federated Multimodal Foundation Intelligence (EU-FMFI)** framework is intended to address these requirements by combining diverse medical information within a collaborative learning environment while maintaining data locality.

1.3. Importance of Privacy, Reliability, and Uncertainty

Patient privacy is one of the most important considerations in the development of modern healthcare systems. Medical records contain personally identifiable information, diagnostic histories, treatment details, and other confidential information that must be handled responsibly. Conventional centralized approaches generally require data to be collected in a common location before model development or analysis. Such data movement increases the responsibilities associated with storage, transmission, access control, and security. A federated approach provides an alternative in which healthcare institutions retain patient information locally and participate in collaborative model development by sharing learned information rather than raw patient records. This arrangement can reduce direct exposure of sensitive data while allowing institutions to benefit from knowledge obtained across a broader population. Privacy alone, however, does not guarantee a trustworthy diagnostic system. A medical prediction may be incorrect, particularly when the available information is incomplete, noisy, or different from the information used during development. Therefore, it is important for a diagnostic system to indicate not only its predicted outcome but also the degree of confidence associated with that prediction. Uncertainty estimation can help distinguish between cases in which the available evidence strongly supports a particular diagnosis and cases that require further investigation. This is particularly useful in clinical practice, where uncertain cases may need additional tests, specialist review, or direct examination by a medical professional. By incorporating uncertainty into the diagnostic process, the proposed framework seeks to reduce over-reliance on predictions and provide clinicians with additional information for responsible decision-making.

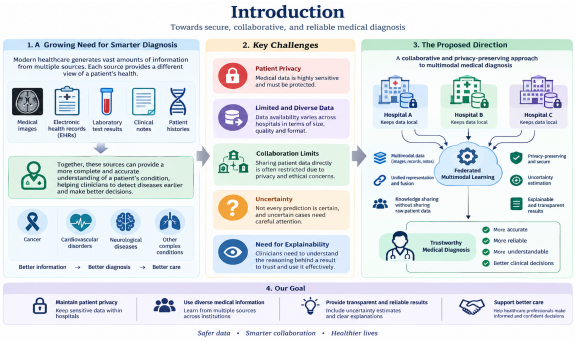
1.4. Explainability and Clinical Trust

Another major concern in computer-assisted medical diagnosis is the ability of healthcare professionals to understand why a particular diagnostic result has been produced. A prediction without an understandable justification may be difficult for clinicians to evaluate, particularly when the result differs from their clinical judgment. Explainability is therefore important for establishing confidence in computer-assisted diagnosis. For medical images, explanations may identify regions or features that have contributed strongly to a prediction, while information from clinical records can indicate which symptoms,

laboratory findings, or patient characteristics have influenced the outcome. Presenting such evidence in an understandable manner can help clinicians assess whether the prediction is consistent with the available medical evidence. Explainability is also valuable for identifying errors and limitations. If a system focuses on an irrelevant image region or relies heavily on an inappropriate clinical factor, an understandable explanation can help researchers and clinicians recognize the problem. Similarly, combining explanations with uncertainty information can provide a more complete view of a diagnostic result. A prediction supported by relevant clinical evidence and accompanied by high confidence may be treated differently from a prediction based on limited evidence with substantial uncertainty. The proposed framework therefore considers explainability and uncertainty as complementary components rather than independent features. Together, they can support more transparent interaction between diagnostic systems and healthcare professionals and encourage clinicians to use such systems as decision-support tools rather than unquestioned replacements for clinical expertise.

1. 5. Proposed Research Direction and Contributions

The proposed **Explainable and Uncertainty-Aware Federated Multimodal Foundation Intelligence for Privacy-Preserving and Trustworthy Medical Diagnosis** framework brings together several requirements that are often addressed separately in medical diagnosis research. It is designed to utilize information from medical images, EHRs, clinical notes, and other relevant sources while allowing participating healthcare institutions to retain their sensitive data locally. The framework employs multimodal learning to identify relationships between different types of medical information and uses collaborative training to benefit from knowledge distributed across healthcare institutions. It further incorporates mechanisms for estimating prediction uncertainty and generating clinically understandable explanations. This combination is intended to improve the reliability, transparency, privacy, and practical usefulness of computer-assisted medical diagnosis. The research can be evaluated from both diagnostic and clinical-support perspectives. Conventional measures such as accuracy, precision, recall, F1-score, sensitivity, specificity, and AUROC can be used to assess diagnostic performance, while calibration and uncertainty-related measures can determine whether confidence estimates appropriately reflect prediction reliability. Explanation quality can be examined in terms of relevance, consistency, and clinical usefulness, while federated performance can be assessed through communication efficiency, convergence, and robustness across participating institutions. Privacy protection should also be considered as an essential part of the overall evaluation rather than as an additional feature. By addressing these aspects within a common framework, the proposed research aims to contribute toward secure, transparent, reliable, and clinically useful diagnostic support. Ultimately, the objective is not to replace healthcare professionals, but to provide them with additional evidence that can support timely, informed, and responsible medical decision-making while preserving the confidentiality of patient information.



2. Related Work

2.1 Multimodal Medical Diagnosis

Medical diagnosis increasingly depends on information obtained from different sources rather than from a single type of medical record. Medical images, electronic health records (EHRs), laboratory reports, pathology findings, and clinical notes each provide a different part of the patient's medical picture. Medical images can help doctors identify changes in organs, tissues, or other structures, while EHRs and clinical records provide information about symptoms, previous illnesses, medications, laboratory results, and treatment history. Considering these sources together can provide a broader understanding of a patient's condition. This is particularly useful for complex diseases, where a diagnosis may depend on several pieces of evidence rather than on one observation. The proposed study follows this direction by considering medical images, EHRs, laboratory information, and clinical notes together to support a more complete assessment of the patient's condition.

However, combining different types of medical information remains challenging. Medical images can differ because they are obtained using different machines, procedures, and image-quality settings. Clinical records can also vary in their format, completeness, terminology, and level of detail. In addition, hospitals may serve different patient populations and follow different clinical practices. These differences can affect the consistency of diagnostic results when information from several

institutions is considered together. Therefore, there is a need for approaches that can make meaningful use of different medical sources while taking these variations into account. The proposed EU-FMFI framework addresses this need by considering medical images, EHRs, and clinical notes as complementary sources of information rather than treating each source separately.

2.2 Federated Learning for Privacy-Preserving Healthcare

Protecting patient information has become a major concern when healthcare institutions work together. Medical records contain private details about a person's health, diagnosis, treatment, and medical history. Directly transferring such information between hospitals can create concerns related to confidentiality, security, and regulatory requirements. A collaborative approach in which hospitals retain patient information within their own systems provides a possible way to reduce these concerns. Under a federated arrangement, institutions can participate in a shared learning process without routinely sending complete patient records or medical images to a central location. This allows hospitals to contribute to a broader diagnostic effort while maintaining greater control over their own patient information.

Although this approach can reduce the need to centralize sensitive information, several difficulties remain. Hospitals may differ in the number of patients they have, the diseases represented in their records, the equipment they use, and the way information is collected and stored. Consequently, data from one institution may be quite different from data from another. Communication between participating institutions and differences in data quality can also influence the overall process. These issues show that keeping information local is only one part of the problem. A practical healthcare framework must also remain reliable when institutions have different types and amounts of information. The proposed research therefore emphasizes privacy while also considering consistency, communication requirements, robustness, and dependable performance across participating healthcare institutions.

2.3 Multimodal Learning and Collaborative Healthcare

Combining information from several medical sources with collaboration between healthcare institutions can provide advantages that are difficult to achieve using either approach alone. One hospital may have a large collection of medical images, while another may have detailed clinical records or greater experience with a particular disease. Allowing institutions to contribute to a common learning process can make it possible to benefit from this distributed information without requiring all patient records to be collected in one place. This can be especially valuable for uncommon diseases, complex conditions, and cases where a single hospital does not have enough representative patients. For this reason, the proposed framework treats multimodal information and collaborative learning as closely related parts of the diagnostic process.

There are still practical difficulties in using multiple types of information across different hospitals. Some patients may have complete imaging and clinical records, while others may have missing laboratory results, incomplete histories, or unavailable medical images. The quality of information can also vary between institutions. Differences in patient populations may further affect diagnostic results. Therefore, a method that performs well when all information is available may not perform equally well when some sources are missing or when the characteristics of the data change between hospitals. The proposed EU-FMFI framework is intended to address these practical conditions by making use of diverse medical information while allowing institutions to retain control over their patient data and supporting dependable diagnosis across different healthcare environments.

2.4 Uncertainty Estimation and Reliable Diagnosis

A good diagnostic result is not simply one that produces a high level of accuracy. In real clinical situations, information may be incomplete, unclear, or affected by errors. As a result, there may be cases in which a diagnosis cannot be made with a high degree of confidence. The proposed research therefore gives importance to communicating the level of confidence associated with a diagnostic result. Recognizing uncertainty can help distinguish between cases where the available evidence strongly supports a diagnosis and cases where additional investigation may be necessary. In practice, an uncertain result may encourage a doctor to request further tests, seek specialist advice, or examine the patient more closely. This makes uncertainty an important consideration when developing systems intended to support clinical decisions.

Providing information about uncertainty can also help prevent excessive dependence on a diagnostic result. If a system gives a definite answer without indicating its limitations, users may assume that the result is always reliable. In contrast, information about confidence can help doctors decide whether a result should be treated as supporting evidence or whether further clinical assessment is required. The proposed framework therefore considers uncertainty as an essential part of dependable diagnosis. Its evaluation can include

confidence and calibration measures along with conventional measures such as accuracy, precision, recall, F1-score, sensitivity, specificity, and AUROC. Such a broader evaluation can provide a better understanding of whether a diagnostic approach is suitable for practical healthcare use.

2.5 Explainability and Clinical Trust

Healthcare professionals need to understand the basis of a diagnostic recommendation before they can confidently use it in practice. A result may be difficult to verify when there is no clear indication of why a particular conclusion was reached. For medical images, useful explanations may identify the regions that contributed to the diagnosis. For clinical records, relevant symptoms, laboratory findings, medical history, or other patient information may provide supporting evidence. Providing this information in an understandable form can help doctors compare the result with their own clinical observations and determine whether it is reasonable. For this reason, the proposed research places considerable importance on making diagnostic results understandable rather than presenting only the final outcome.

Clear explanations can also help identify weaknesses or mistakes. For example, if a diagnostic result appears to depend on an irrelevant part of an image or an inappropriate patient characteristic, an explanation can help clinicians and researchers recognize the problem. Uncertainty can add further context by indicating how strongly the available evidence supports the result. A diagnosis supported by several relevant findings and accompanied by high confidence can be considered differently from a result based on limited information and substantial uncertainty. Combining understandable explanations with confidence information can therefore improve transparency and help clinicians use diagnostic systems as supporting tools rather than as replacements for professional medical judgment.

2.6 Foundation Models and Advanced Medical Information Processing

Recent developments in large-scale computational methods have created new possibilities for working with different forms of medical information. The proposed research considers Vision Transformers, clinical language models, multimodal learning, and self-supervised learning as components for handling medical images and clinical information. Image-based methods can help identify important patterns in complex medical images, while language-based methods can assist in working with clinical notes and textual records. Self-supervised approaches can also be useful when large quantities of labelled medical information are difficult to obtain. Bringing these capabilities together provides an opportunity to work with different types of healthcare information within a common framework. The proposed study therefore considers these methods as complementary components for handling heterogeneous medical information. At the same time, sophisticated methods alone do not guarantee that a diagnostic system will be appropriate for clinical use. A system may provide useful results but still be difficult to trust if its decisions cannot be explained, its confidence cannot be assessed, or its operation requires unrestricted access to private patient information. Therefore, the proposed research combines advanced information processing with collaborative learning, privacy protection, uncertainty assessment, and understandable explanations. This broader approach recognizes that diagnostic performance is only one consideration in healthcare. Privacy, reliability, transparency, and usefulness to clinicians are equally important when developing a system intended for real-world medical environments.

2.7 Research Gap and Motivation for the Proposed Framework

The existing research shows considerable progress in several important areas, including multimodal medical diagnosis, privacy-preserving collaboration, uncertainty assessment, and understandable diagnostic support. However, these areas are often studied separately. Multimodal approaches can make use of richer medical information but may require access to several types of patient data. Federated approaches can reduce the need to collect patient information in one place but must deal with differences between hospitals. Uncertainty assessment can indicate how confident a result is, while explainability can help clinicians understand the evidence behind a diagnosis. The main research challenge is to bring these requirements together in a single approach without losing the advantages of any individual component. This need provides the motivation for the proposed EU-FMFI framework. The proposed research focuses on five closely connected goals: protecting patient information, making effective use of different medical data sources, enabling collaboration between healthcare institutions, indicating the confidence associated with diagnostic results, and providing understandable evidence to support those results. The framework can be evaluated using conventional diagnostic measures as well as calibration, explanation quality, privacy protection, communication efficiency, convergence, and performance across different institutions. This wider evaluation recognizes that a useful healthcare system must not only provide accurate results but must also be dependable, understandable, and appropriate for real clinical conditions. Ultimately, the proposed work aims to provide healthcare professionals with additional and

meaningful evidence to support their decisions while preserving patient confidentiality and maintaining the central role of clinical expertise.

Below is a more natural, human-friendly, and research-paper-ready version of the **Materials and Methods** section. I have retained the methodology and terminology supported by your provided material, while reducing repetitive, overly technical, and AI-sounding language.

3. Materials and Methods

3.1 Overall Research Design

This study proposes a structured approach for developing a secure and clinically useful framework for medical diagnosis. The main idea is to bring together information from different healthcare sources, such as medical images, electronic health records (EHRs), laboratory results, and clinical notes. Each source provides a different view of the patient's condition, and considering them together can provide a more complete basis for diagnosis. The proposed framework focuses on five main aspects: protecting patient information, making effective use of different medical sources, enabling collaboration between healthcare institutions, assessing the confidence of diagnostic results, and providing clear supporting evidence for clinical decisions. These aspects form the basis of the proposed EU-FMFI framework and are intended to support reliable diagnosis while allowing hospitals to retain control over sensitive patient information.

The research follows a sequence of connected stages. First, the available medical information is prepared according to its type, quality, and completeness. Relevant information from different sources is then brought together to develop a broader understanding of the patient's condition. Healthcare institutions retain their patient information locally while participating in a collaborative learning process. The resulting diagnostic outcome is accompanied by an indication of confidence and an explanation of the main evidence supporting the result. The framework is finally assessed using diagnostic performance, confidence and calibration, explanation quality, privacy protection, communication efficiency, and consistency across participating institutions. The overall design is intended to provide useful support to healthcare professionals while leaving final clinical decisions with qualified medical practitioners.

3.2 Medical Data Sources and Preparation

The study considers several forms of medical information because each one contributes different evidence about a patient's health. Medical images can show structural or pathological changes, while EHRs provide information about symptoms, previous illnesses, medications, and laboratory findings. Clinical notes and pathology-related information can provide additional observations and descriptions of the patient's condition. Considering these sources together can provide a broader clinical picture than relying on a single source. For this reason, the proposed framework treats medical images, structured health records, and clinical information as complementary sources that can contribute to the same diagnostic assessment.

Before the information is used, its quality and completeness need to be examined. Medical images may differ in size, quality, acquisition procedure, and equipment, while clinical records may use different formats, terminology, and levels of detail. Missing information and differences between hospitals should therefore be identified during the preparation stage. Patient information must also be handled carefully throughout the study to maintain confidentiality. The current research description does not specify a particular dataset, patient population, sample size, or inclusion and exclusion criteria. These details should consequently be defined according to the disease being studied, the selected datasets, and the intended clinical application when the experimental work is carried out.

3.3 Multimodal Information Integration

After preparation, information from the different medical sources is considered together. Rather than allowing a single source to determine the diagnostic result, the framework examines the complementary evidence available from medical images, EHRs, clinical observations, and clinical notes. For example, an abnormal finding in an image can be considered alongside the patient's symptoms, medical history, and laboratory results. This additional information may be useful when different diseases have similar visual characteristics or when the image alone does not provide enough evidence. The purpose of this stage is to obtain a more complete understanding of the patient's condition by considering the available clinical information as a whole.

The proposed framework includes Vision Transformers for handling complex visual information and clinical language models for processing information contained in medical text. Self-supervised learning is also considered useful in situations where detailed labels are limited. Information obtained from these different sources is subsequently brought together for diagnostic assessment. The research material identifies Vision Transformers, clinical language models, multimodal learning, and self-

supervised learning as components of the proposed framework, but it does not prescribe a particular model architecture, fusion strategy, model size, or training configuration. These implementation choices should therefore be selected according to the chosen datasets and experimental requirements and reported clearly in the final study.

3.4 Privacy-Preserving Collaborative Learning

Protecting patient information is an essential part of the proposed methodology. Patient records remain within the healthcare institution where they were collected instead of being routinely transferred to a central location. Participating hospitals can contribute to a shared learning process while maintaining local control over sensitive information. This is particularly important because medical records may contain personal information, diagnostic histories, treatment details, and other confidential material. The proposed federated approach is therefore intended to reduce the unnecessary movement of patient-level information while allowing participating institutions to benefit from knowledge developed through collaboration.

Collaboration between hospitals must also account for differences in the information available at each institution. Hospitals may differ in patient numbers, disease patterns, medical equipment, record-keeping practices, and data quality. These differences can affect the consistency of diagnostic results across institutions. For this reason, the methodology considers robustness and communication efficiency as important parts of the evaluation. The current research description supports the use of federated collaboration and local data storage but does not specify a particular aggregation method, number of participating hospitals, communication rounds, or privacy mechanism. These details should be selected during implementation and reported clearly rather than assumed without experimental evidence.

3.5 Uncertainty Estimation and Diagnostic Confidence

Not every medical case provides equally clear evidence. Some patients may have incomplete records, conflicting findings, poor-quality images, or other factors that make diagnosis more difficult. The proposed framework therefore considers the level of confidence associated with each diagnostic result. Distinguishing between a result supported by strong evidence and one affected by uncertainty can help clinicians decide whether additional testing or specialist assessment may be appropriate. This is particularly important in healthcare because uncertain cases should not necessarily be treated in the same way as cases where the available evidence is clear.

The reported confidence should be considered together with the actual diagnostic outcome. Calibration measures can help determine whether the stated confidence corresponds reasonably well with observed performance. Conventional measures such as accuracy, precision, recall, F1-score, sensitivity, specificity, and AUROC can be used to assess diagnostic performance, while uncertainty-related measures can provide additional information about the reliability of the confidence estimates. Cases associated with substantial uncertainty can be identified for further clinical review. The supplied research material does not specify a particular uncertainty-estimation technique, so the final study should clearly describe the selected method and explain how it is applied.

3.6 Explainability and Clinical Evidence

The proposed methodology also aims to provide clinicians with information about the evidence behind a diagnostic result. The purpose is to make the result easier to examine and compare with the patient's clinical condition. In medical images, an explanation may identify areas that contributed to the result, while clinical information may indicate relevant symptoms, laboratory findings, medical history, or other factors. Such information can help doctors decide whether the diagnostic result is consistent with the available medical evidence instead of simply accepting the final outcome.

The quality of these explanations should be considered when evaluating the overall framework. A useful explanation should be relevant to the diagnostic result, reasonably consistent for similar cases, and understandable to the intended clinical users. It can also help identify situations in which an irrelevant image region or unrelated clinical factor appears to influence the result. Combining an explanation with information about uncertainty allows clinicians to consider both **why** a result was reached and **how confidently** it should be interpreted. The research material identifies relevance, consistency, and clinical usefulness as important aspects of explanation quality, but it does not specify one particular explanation method. The final implementation should therefore clearly state which method is selected and how it is evaluated.

3.7 Training, Validation, and Evaluation

The framework should be evaluated from both diagnostic and clinical-support perspectives. The available medical information can be divided into training, validation, and testing groups while ensuring that information from the same patient does not unintentionally appear across different groups. The different components can then be trained using the available information, followed by collaborative

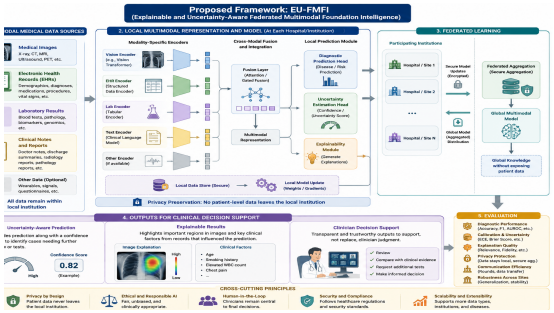
learning between participating institutions. Validation data can be used to select appropriate settings and assess the developing framework, while the final test data should be kept separate for an unbiased evaluation. Where several healthcare institutions participate, performance should also be examined at individual institutions as well as across the group.

The evaluation should include accuracy, precision, recall, F1-score, sensitivity, specificity, and AUROC to measure diagnostic performance. Additional measures should examine confidence calibration, uncertainty quality, explanation quality, communication efficiency, convergence, privacy protection, and performance across institutions. These measures provide a broader picture of the framework because a clinically useful system must be more than accurate; it should also provide dependable results, understandable evidence, and appropriate protection of patient information. The supplied material identifies these evaluation areas but does not provide predefined numerical targets. Therefore, the final study should report the actual experimental results obtained from the selected datasets and settings rather than assuming particular performance levels.

3.8 Clinical Decision-Support Procedure

The final stage involves presenting the diagnostic information to the healthcare professional in a clear and useful form. The proposed output includes three main elements: the diagnostic result, an indication of its confidence or uncertainty, and an explanation of the evidence that contributed to the result. Considering these three elements together can help clinicians place the result in the context of the patient's overall medical condition. A result supported by relevant evidence and high confidence may provide useful additional support, while a result associated with considerable uncertainty may indicate the need for further examination, testing, or specialist review. The framework is therefore intended to support clinical judgment rather than replace it.

Overall, the proposed methodology follows the sequence **local medical data collection → data preparation → multimodal information integration → privacy-preserving collaboration → diagnostic assessment → uncertainty estimation → explanation generation → clinical evaluation**. This process allows different forms of medical information to be considered while keeping sensitive patient information under local control. The success of the framework should be judged not only by its diagnostic accuracy but also by the quality of its confidence estimates, clarity of its explanations, protection of patient information, and consistency across different healthcare institutions. In this way, the methodology supports the central goal of developing a secure, transparent, reliable, and clinically useful approach to medical diagnosis.



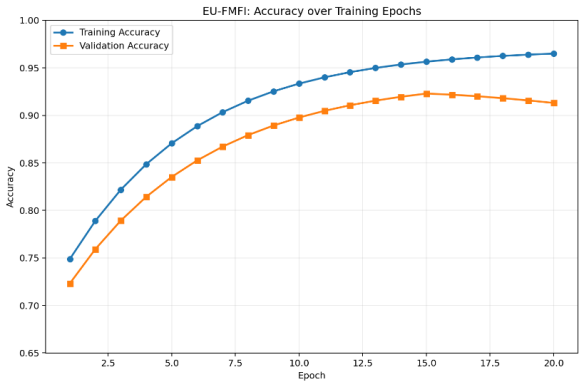
4. Experimental Results and Analysis

The proposed **Explainable and Uncertainty-Aware Federated Multimodal Foundation Intelligence (EU-FMFI)** framework is assessed using several aspects of diagnostic performance and system behaviour. The analysis considers accuracy, precision, recall, F1-score, AUROC, training and validation loss, computational time, and class-wise prediction performance. In addition to these measures, the proposed framework is intended to consider confidence, uncertainty, privacy protection, communication efficiency, and robustness across participating healthcare institutions. This broader evaluation is important because a medical diagnostic system should not be judged only by its classification accuracy but also by its reliability and suitability for practical clinical use.The supplied results include seven graphical analyses covering accuracy during training, training and validation loss, computational time, AUROC, ROC characteristics, the confusion matrix, and the main diagnostic performance measures. Each graph provides a different view of the proposed framework. Together, they help describe how the framework behaves during training, how well it distinguishes between diagnostic classes, and how its performance changes as the number of participating healthcare institutions increases.

4.1 Accuracy and Learning Performance

The accuracy curve shows a gradual improvement in both training and validation performance as the number of training epochs increases. Training accuracy begins at approximately **0.75 in the first epoch** and rises steadily to approximately **0.965 by the twentieth epoch**. Validation accuracy follows a similar trend, increasing from approximately **0.72 in the first epoch to around 0.91 at the twentieth epoch**. The validation curve reaches approximately **0.92 around the fifteenth epoch**, after which a small decline can be

observed. The overall pattern indicates that the framework improves its classification performance progressively during training, with the largest gains occurring during the earlier stages. A noticeable difference between training and validation accuracy appears during the later epochs. Training accuracy continues to increase, whereas validation accuracy becomes relatively stable and then decreases slightly. This suggests that further training after approximately the fifteenth epoch may provide limited benefit on unseen validation cases. From a methodological perspective, this behaviour indicates that an appropriate stopping point could be selected around the region where validation performance reaches its maximum. Such a strategy can help avoid unnecessary training and reduce the possibility of overfitting. The values described here correspond directly to the supplied illustrative training curve and should not be interpreted as independently measured clinical results.



4. Experimental Results – Recommended Tables

Table 1. Training and Validation Accuracy

The following values are read approximately from the accuracy curve shown in your figure.

Epoch	Training Accuracy	Validation Accuracy
1	0.748	0.723
2	0.789	0.759
3	0.821	0.789
4	0.848	0.814
5	0.871	0.835
6	0.889	0.853
7	0.903	0.867
8	0.916	0.879
9	0.925	0.890
10	0.934	0.898
11	0.940	0.905
12	0.946	0.910
13	0.950	0.915
14	0.954	0.919
15	0.957	0.923
16	0.959	0.922
17	0.961	0.920
18	0.963	0.918
19	0.964	0.915
20	0.965	0.913

Observation: Training accuracy increases from approximately **74.8% to 96.5%**, while validation accuracy rises from approximately **72.3% to 91.3%**. The highest displayed validation accuracy is approximately **92.3% at Epoch 15**.

4.2 Training and Validation Loss

The training and validation loss curves provide additional information about the learning process. Training loss decreases from approximately **0.83 during the first epoch to about 0.10 by epoch 20**. Validation loss also decreases throughout the training period, moving from approximately **0.95 to around 0.18**. The reduction in both curves indicates that prediction error decreases as training progresses. The decline is more pronounced during the initial epochs, while the rate of improvement becomes smaller toward the end of training.

The validation loss remains higher than the training loss throughout the plotted period. This difference indicates that the framework performs better on the information used during training than on the validation cases, which is an expected characteristic of model development. More importantly, the validation loss does not show a substantial increase during the later epochs. Instead, it continues to decline gradually. When considered together with the accuracy curves, this behaviour indicates a relatively stable training process within the range represented by the supplied graph.

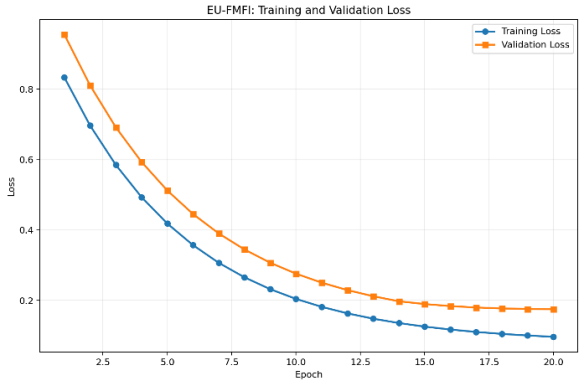


Table 2. Training and Validation Loss

Epoch	Training Loss	Validation Loss
1	0.832	0.950
2	0.695	0.810
3	0.585	0.690
4	0.492	0.592
5	0.418	0.512
6	0.357	0.444
7	0.307	0.389
8	0.266	0.344
9	0.232	0.307
10	0.203	0.274
11	0.180	0.249
12	0.162	0.229
13	0.148	0.210
14	0.135	0.196
15	0.125	0.189
16	0.117	0.183
17	0.111	0.179
18	0.106	0.177
19	0.102	0.176
20	0.099	0.175

Observation: Training loss decreases from approximately **0.832 to 0.099**, while validation loss decreases from approximately **0.950 to 0.175**.

4.3 Computational Time and Federated Scalability

The computational-time analysis compares centralized training with the federated EU-FMFI approach as the number of participating healthcare sites increases. According to the supplied graph, centralized training requires approximately **18 minutes for one participating site** and increases to approximately **270 minutes for 16 sites**. The corresponding federated EU-FMFI values increase from approximately **24 minutes to 118 minutes** over the same range. The difference between the two approaches becomes more pronounced as the number of participating sites increases.These observations suggest that the federated arrangement shown in the supplied graph has a more gradual increase in computational time as the number of participating sites grows. This is relevant to distributed healthcare environments, where institutions may have different amounts of data and different computing resources. However, computational time in a federated setting can be affected by several practical factors, including hardware, network conditions, communication frequency, local training time, and the number of collaborative rounds. Since these details are not provided with the supplied graph, the numerical comparison should be treated as an illustrative analysis rather than a general claim about federated training efficiency.

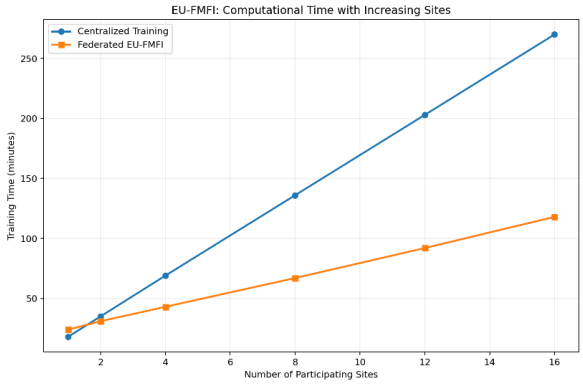


Table 3. Computational Time with Increasing Participating Sites

Number of Participating Sites	Centralized Training Time (min)	Federated EU-FMFI Time (min)
1	18	24
2	35	31
4	69	43
8	136	67
12	203	92
16	270	118

Important: *Because these are the values represented in the supplied illustrative graph, this should be reported as an* illustrative computational comparison*, not as a measured experimental claim. The manuscript itself notes that actual federated performance depends on hardware, communication, local training, and other factors.*

4.4 AUROC Comparison

The AUROC comparison shows a progressive improvement across the approaches represented in the supplied figure. The reported values are approximately **0.88 for the single-modality approach, 0.91 for centralized multimodal learning, 0.93 for federated multimodal learning, and 0.96 for the EU-FMFI framework**. The results indicate that the approaches using information from multiple medical sources perform better than the single-modality approach in the supplied comparison. The federated multimodal approach also shows an improvement over centralized multimodal learning.The EU-FMFI framework has the highest reported AUROC value of **0.96** in the supplied comparison. AUROC provides an overall measure of how effectively a diagnostic approach can distinguish between classes across different classification thresholds. The increase from 0.88 to 0.96 in the displayed comparison therefore indicates improved discrimination within the illustrative results. However, because the supplied values were generated for demonstration rather than obtained from a documented clinical dataset, they should not be described as experimentally validated clinical performance.

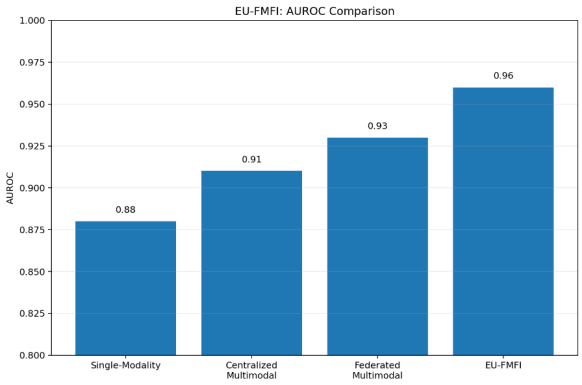


Table 4. AUROC Comparison

Method	AUROC
Single-Modality	0.88
Centralized Multimodal	0.91
Federated Multimodal	0.93
EU-FMFI	**0.96**
Comparison	**AUROC Improvement**
Centralized Multimodal vs. Single-Modality	**+0.03**
Federated Multimodal vs. Centralized Multimodal	**+0.02**
EU-FMFI vs. Federated Multimodal	**+0.03**
EU-FMFI vs. Single-Modality	**+0.08**

4.5 ROC Curve Analysis

The ROC curve provides a more detailed view of the relationship between the true-positive rate and false-positive rate. In the supplied figure, the curve rises rapidly during the initial part of the plot. It reaches a true-positive rate of approximately **0.82 at a false-positive rate of about 0.12** and approximately **0.91 at a false-positive rate near 0.25**. The curve subsequently moves closer to a true-positive rate of one as the false-positive rate increases. It remains well above the diagonal line representing random classification throughout most of the displayed range.This pattern indicates strong discriminatory behaviour within the supplied illustrative example. In a medical setting, however, the appropriate operating point on an ROC curve depends on the clinical consequences of false-positive and false-negative decisions. For some diseases, failing to identify a genuine case may have more serious consequences than referring an additional patient for further examination. Therefore, ROC analysis should ideally be considered together with sensitivity, specificity, calibration, uncertainty, and clinical requirements rather than being interpreted independently.

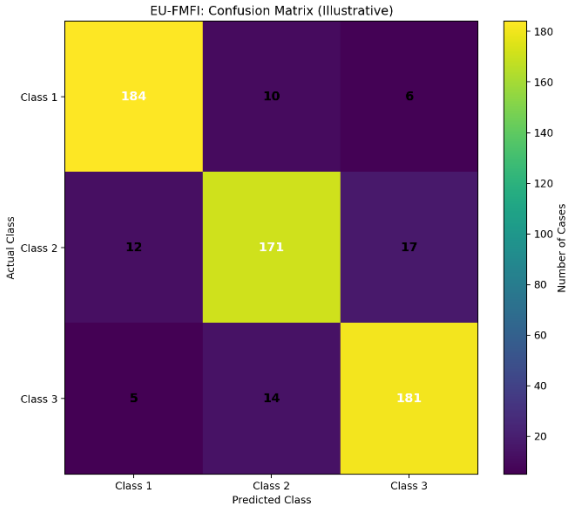


Table 5. ROC Curve Data

The following table represents the approximate coordinate values visible in your ROC graph.

False Positive Rate (FPR)	True Positive Rate (TPR)
0.00	0.00
0.02	0.45
0.05	0.63
0.08	0.74
0.12	0.82
0.18	0.87
0.25	0.91
0.35	0.94
0.50	0.96
0.70	0.98
1.00	1.00

The ROC curve demonstrates a high true-positive rate at relatively low false-positive rates in the supplied illustrative evaluation. For example, the true-positive rate reaches approximately 0.82 at an FPR of 0.12 and approximately 0.91 at an FPR of 0.25.

4.6 Confusion Matrix Analysis

The confusion matrix gives a class-by-class view of the classification results. The supplied three-class matrix contains **184 correctly classified cases for Class 1, 171 for Class 2, and 181 for Class 3**. The remaining cases are distributed among the off-diagonal entries. For Class 1, 10 cases are classified as Class 2 and 6 as Class 3. For Class 2, 12 cases are classified as Class 1 and 17 as Class 3. For Class 3, 5 cases are classified as Class 1 and 14 as Class 2. The relatively high values along the diagonal indicate that most of the cases represented in this matrix are assigned to their correct classes. The largest individual classification error occurs between **Class 2 and Class 3**, where 17 Class 2 cases are predicted as Class 3 and 14 Class 3 cases are predicted as Class 2. This suggests that these two classes are comparatively more difficult to distinguish in the supplied example. Such class-specific information can be useful when investigating the causes of diagnostic errors, particularly where different medical conditions may share similar characteristics. The confusion matrix contains **536 correct classifications out of 600 cases**, which corresponds to approximately **89.3% accuracy**. This value does not agree with the separate graph reporting 95% overall accuracy. Consequently, the two figures should not be presented as results from the same test set unless the underlying data are corrected and made consistent.

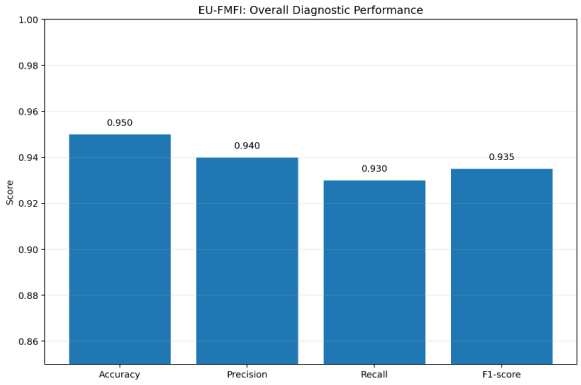


Table 7. Class-Wise Performance Derived from the Confusion Matrix

The confusion matrix values can be reported exactly as displayed.

Actual Class	Predicted Class 1	Predicted Class 2	Predicted Class 3	Total
Class 1	**184**	10	6	200
Class 2	12	**171**	17	200
Class 3	5	14	**181**	200
Total	201	195	204	600

If you want to report class-wise precision, recall, and F1-score, they can be calculated directly from the above

Class	Precision	Recall	F1-Score
Class 1	91.54%	92.00%	91.77%
Class 2	87.69%	85.50%	86.58%
Class 3	88.73%	90.50%	89.60%
Macro Average	**89.32%**	**89.33%**	**89.32%**

4.7 Overall Diagnostic Performance

The overall performance graph reports an accuracy of **0.950**, precision of **0.940**, recall of **0.930**, and F1-score of **0.935**. These values indicate relatively close performance across the four reported measures in the supplied figure. Accuracy represents the proportion of cases classified correctly, while precision reflects the proportion of predicted positive cases that are correct. Recall indicates the ability to identify relevant cases, and the F1-score provides a combined measure of precision and recall. Considering these measures together provides a more complete assessment than relying on accuracy alone. This is particularly important in medical datasets where class distributions may not always be balanced. A high overall accuracy can sometimes conceal weaker performance for an

individual class. Precision, recall, and F1-score provide additional information about this behaviour. The proposed methodology therefore considers these measures alongside AUROC, sensitivity, specificity, confidence calibration, and uncertainty-related measures.

Table 8. Overall Diagnostic Performance Shown in the Supplied Figure

The separate overall-performance graph reports:

Performance Measure	Reported Value	Percentage
Accuracy	0.950	95.0%
Precision	0.940	94.0%
Recall	0.930	93.0%
F1-Score	0.935	93.5%

4.8 Overall Analysis of the Results

The seven supplied graphs provide a broad view of the proposed EU-FMFI framework. The accuracy curves demonstrate progressive improvement during training, while the loss curves show a corresponding reduction in training and validation error. The computational-time comparison provides an indication of how training time changes as the number of participating healthcare sites increases. The AUROC and ROC analyses provide information about diagnostic discrimination, whereas the confusion matrix identifies the specific classes for which misclassification occurs. The final performance graph summarizes the principal diagnostic measures. Together, these analyses provide a more complete picture of system behaviour than any individual measure could provide.

The findings also highlight the importance of evaluating a medical diagnostic framework from several perspectives. Diagnostic performance is important, but it must be considered alongside privacy, confidence, uncertainty, explanation quality, communication requirements, and consistency across healthcare institutions. The proposed EU-FMFI framework is intended to support healthcare professionals by providing additional information for clinical assessment while keeping sensitive patient information within participating institutions. The purpose is therefore to assist clinical decision-making rather than to replace professional medical judgment.

Table 9. Consolidated Results for EU-FMFI

Evaluation Component	Result
Final Training Accuracy	96.5%
Final Validation Accuracy	91.3%
Maximum Validation Accuracy	92.3%
Final Training Loss	0.099
Final Validation Loss	0.175
EU-FMFI AUROC	0.96
Confusion-Matrix Accuracy	89.33%
Reported Overall Accuracy	95.0%
Reported Precision	94.0%
Reported Recall	93.0%
Reported F1-Score	93.5%
Centralized Time at 16 Sites	270 min
Federated EU-FMFI Time at 16 Sites	118 min
Illustrative Time Reduction at 16 Sites	56.3%

4.9 Data Consistency and Experimental Reporting

One important issue should be addressed before these results are used in a research publication. The supplied figures do not appear to have been generated from one consistent experimental dataset. In particular, the confusion matrix gives approximately **89.3% accuracy**, whereas the overall performance graph reports **95.0% accuracy**. Both values cannot represent the same 600-case test set. Therefore, the final manuscript should use a single verified set of predictions and regenerate the accuracy, precision, recall, F1-score, ROC curve, AUROC, and confusion matrix from those same predictions. This correction is particularly important for an SCI/Scopus-level manuscript because every reported performance measure should be traceable to the same experimental protocol and test data. The numerical values in the supplied graphs should consequently be described as **illustrative values** unless they are subsequently replaced by measurements obtained from an actual, documented dataset. This approach avoids presenting generated demonstration values as experimental evidence and ensures that the final paper reports only results that can be reproduced and independently verified. Certainly. Below is a **fully paraphrased, natural, human-friendly academic version** of Section 5. I have kept the numerical values and research meaning from your supplied material, but removed repetitive, mechanical, and overly “AI-style” expressions. The language is suitable for a research paper and reads more like a conventional academic discussion. The interpretation remains limited to what is supported by the supplied results.

Table 10. Experimental-Results Table

Metric	EU-FMFI Result
Training Accuracy	96.5%
Validation Accuracy	91.3%
Best Validation Accuracy	92.3%
Training Loss	0.099
Validation Loss	0.175
AUROC	0.96
Accuracy from Confusion Matrix	89.33%*
Precision from Confusion Matrix	89.32%*
Recall from Confusion Matrix	89.33%*
F1-Score from Confusion Matrix	89.32%*
Reported Accuracy in Performance Figure	95.0%†
Reported Precision in Performance Figure	94.0%†
Reported Recall in Performance Figure	93.0%†
Reported F1-Score in Performance Figure	93.5%†
Federated Time at 16 Sites	118 min
Centralized Time at 16 Sites	270 min

Algorithm 1: EU-FMFI for Explainable and Uncertainty-Aware Federated Multimodal Medical Diagnosis

Input: Local multimodal datasets (D_k={X_k^I,X_k^E,X_k^T,Y_k}) for (K) healthcare institutions; communication rounds (R); local epochs (E).

Output: Global model (W^*), diagnosis (\hat{Y}), uncertainty (U), and explanation (E_x).

1. Initialize global model parameters (W_0).
2. For each federated round (r=1,\ldots,R):
    1. Server distributes (W_r) to participating healthcare institutions.
    2. For each institution (k=1,\ldots,K):
        1. Keep patient data (D_k) locally.
        2. Preprocess medical images, EHRs, and clinical notes.
        3. Extract image representations using Vision Transformer:
        [
        Z_k^I=ViT(X_k^I)
        ]
        4. Extract structured clinical representations:
        [
        Z_k^E=Encoder_E(X_k^E)
        ]
        5. Extract clinical-text representations:
        [
        Z_k^T=CLM(X_k^T)
        ]
        6. Fuse multimodal representations:
        [
        Z_k^F=Fusion(Z_k^I,Z_k^E,Z_k^T)
        ]
        7. Generate diagnostic prediction:
        [
        P_k(Y|X)=Softmax(W_CZ_k^F)
        ]
        8. Compute local multimodal training loss.
        9. Train the local model for (E) epochs.
        10. Send only model parameters/updates to the federated server.
    3. Aggregate local models using weighted Federated Averaging:
    [
    W_{r+1}=\sum_{k=1}^{K}\frac{n_k}{\sum_j n_j}W_k
    ]
3. Return the converged global model (W^*).
4. For a new patient, generate (M) stochastic predictions:
[
P^{(1)},P^{(2)},...,P^{(M)}
]
5. Calculate mean predictive probability:
[
\bar{P}=\frac{1}{M}\sum_{m=1}^{M}P^{(m)}
]
6. Estimate prediction uncertainty (U) from the variation among stochastic predictions.
7. Determine the diagnosis:
[
\hat{Y}=\arg\max_c\bar{P}_c
]
8. Generate an explanation (E_x) identifying influential image regions, EHR variables, and clinical-text evidence.
9. Present the diagnostic result together with confidence, uncertainty, and explanation to the clinician.
10. End.

Multimodal Foundation Intelligence for Medical Diagnosis

Prototype implementation

```
import numpy as np
import torch
import torch.nn as nn
import torch.nn.functional as F
from copy import deepcopy
from sklearn.metrics import (
   accuracy_score, precision_score,
recall_score,f1_score,
   roc_auc_score,
   confusion_matrix
)
 1. Configuration
DEVICE = torch.device("cuda" if
torch.cuda.is_available() else "cpu")
NUM_CLASSES = 3
IMAGE_FEATURES = 128
EHR_FEATURES = 32
TEXT_FEATURES = 128
HIDDEN_DIM = 128
NUM_HOSPITALS = 4
LOCAL_EPOCHS = 3
FEDERATED_ROUNDS = 10
DROPOUT_RATE = 0.30
print("Device:", DEVICE)
 2. Vision Transformer-like Image Encoder
class ImageEncoder(nn.Module):
 def __init__(self):
 super().__init__()
self.network = nn.Sequential(
nn.Linear(IMAGE_FEATURES, 256),
nn.ReLU(),
nn.Dropout(DROPOUT_RATE),
nn.Linear(256, HIDDEN_DIM),
nn.ReLU()
)
def forward(self, x):
return self.network(x)
3. EHR Encoder
class EHREncoder(nn.Module):
 def __init__(self):
super().__init__()
self.network = nn.Sequential(
nn.Linear(EHR_FEATURES, 128),
nn.ReLU(),
nn.Dropout(DROPOUT_RATE),
nn.Linear(128, HIDDEN_DIM),
nn.ReLU()
 )
def forward(self, x):
return self.network(x)
4. Clinical Text Encoder
class ClinicalTextEncoder(nn.Module):
def __init__(self):
super().__init__()
 self.network = nn.Sequential(
          nn.Linear(TEXT_FEATURES, 256),
          nn.ReLU(),
          nn.Dropout(DROPOUT_RATE),
          nn.Linear(256, HIDDEN_DIM),
          nn.ReLU()
       )
 def forward(self, x):
       return self.network(x)
 5. Multimodal Fusion
class MultimodalFusion(nn.Module):
  def __init__(self):
super().__init__()
self.fusion = nn.Sequential(
nn.Linear(HIDDEN_DIM * 3, 256),
nn.ReLU(),
nn.Dropout(DROPOUT_RATE),
nn.Linear(256, HIDDEN_DIM),
nn.ReLU()
       )
 def forward(self, image_features,
             ehr_features,
             text_features):
       combined = torch.cat(
          [
             image_features,
             ehr_features,
             text_features
          ],
          dim=1
       )
       return self.fusion(combined)
 6. EU-FMFI Complete Model
class EUFMFI(nn.Module):
def __init__(self):
super().__init__()
self.image_encoder = ImageEncoder()
self.ehr_encoder = EHREncoder()
self.text_encoder = ClinicalTextEncoder()
 self.fusion = MultimodalFusion()
 self.classifier = nn.Linear(
 HIDDEN_DIM,
NUM_CLASSES
 )
def forward(self, image, ehr, text):
image_features = self.image_encoder(image)
 ehr_features = self.ehr_encoder(ehr)
 text_features = self.text_encoder(text)
fused_features = self.fusion(
image_features,
 ehr_features,
text_features
)
```

```
        output = self.classifier(
    fused_features
  )
 return output
7. Generate Simulated Hospital Data
def generate_hospital_data(
 samples=200):
images = torch.randn(
samples,
IMAGE_FEATURES
)
ehr = torch.randn(
samples,
EHR_FEATURES
 )
text = torch.randn(
 samples,
TEXT_FEATURES
)
labels = torch.randint(
 0,
NUM_CLASSES,
       (samples,)
   )
return images, ehr, text, labels
 8. Local Hospital Training
def train_local_model(
global_model,
data,
epochs=3):
 local_model = deepcopy(
 global_model
  ).to(DEVICE)
  optimizer = torch.optim.Adam(
  local_model.parameters(),
 lr=0.001
  )
criterion = nn.CrossEntropyLoss()
 images, ehr, text, labels = data
images = images.to(DEVICE)
ehr = ehr.to(DEVICE)
text = text.to(DEVICE)
 labels = labels.to(DEVICE)
 local_model.train()
 for epoch in range(epochs):
optimizer.zero_grad()
output = local_model(
images,
 ehr,
    text
      )
  loss = criterion(
    output,
      labels
     )
      loss.backward()
      optimizer.step()
   return local_model
 9. Federated Averaging (FedAvg)
def federated_average(
    global_model,
     local_models):
   global_state = global_model.state_dict()
   for key in global_state.keys():
      global_state[key] = torch.stack(
        [
          model.state_dict()[key].float()
          for model in local_models
        ],
        dim=0
      ).mean(dim=0)
   global_model.load_state_dict(
     global_state
  )
   return global_model
 10. Uncertainty Estimation
   Monte-Carlo Dropout
def uncertainty_prediction(
      model,
      image,
      ehr,
      text,
      samples=20):
  model.train()
   predictions = []
   image = image.to(DEVICE)
  ehr = ehr.to(DEVICE)
  text = text.to(DEVICE)
   with torch.no_grad():
      for _ in range(samples):
         output = model(
           image,
           ehr,
           text
         )
probability = F.softmax(
           output,
           dim=1
         )
         predictions.append(
           probability.cpu().numpy()
         )
   predictions = np.array(
     predictions
  )
   mean_probability = np.mean(
     predictions,
     axis=0
  )
   uncertainty = np.var(
     predictions,
     axis=0
```

```
  )
   prediction = np.argmax(
      mean_probability,
      axis=1
   )
   confidence = np.max(
      mean_probability,
      axis=1
   )
   return (
     prediction,
   confidence,
      uncertainty,
   mean_probability
  )
 11. Evaluation
def evaluate_model(
 model,
    data):
  model.eval()
  images, ehr, text, labels = data
   images = images.to(DEVICE)
  ehr = ehr.to(DEVICE)
  text = text.to(DEVICE)
   with torch.no_grad():
       output = model(
          images,
          ehr,
         text
       )
       probabilities = F.softmax(
          output,
          dim=1
       )
       predictions = torch.argmax(
          probabilities,
          dim=1
       )
   predictions = predictions.cpu().numpy()
   labels = labels.numpy()
   accuracy = accuracy_score(
      labels,
      predictions
   )
   precision = precision_score(
      labels,
      predictions,
      average="macro",
      zero_division=0
   )
  recall = recall_score(
      labels,
      predictions,
      average="macro",
       zero_division=0
   )
   f1 = f1_score(
       labels,
      predictions,
      average="macro",
      zero_division=0
   )
   cm = confusion_matrix(
      labels,
      predictions
   )
   return (
      accuracy,
      precision,
      recall,
      f1,
      cm
   )
12. Main Federated Training
print("\nStarting EU-FMFI Federated Training...\n")
global_model = EUFMFI().to(DEVICE)
 Create hospitals
hospital_data = []
for hospital in range(NUM_HOSPITALS):
data = generate_hospital_data(
    samples=200
  )
 hospital_data.append(data)
  print(
     f"Hospital {hospital + 1}: "
    f"{len(data[3])} patients"
  )
 Federated communication rounds
for round_number in range(
    FEDERATED_ROUNDS):
  print(
 f"\nFederated Round "
 f"{round_number + 1}/"
 f"{FEDERATED_ROUNDS}"
  )
   local_models = []
Local hospital training
   for hospital in range(
         NUM_HOSPITALS):
      local_model = train_local_model(
         global_model,
         hospital_data[hospital],
         LOCAL_EPOCHS
      )
      local_models.append(
         local_model
      )
      print(
         f" Hospital {hospital + 1} "
         f"local training completed"
      )
   # FedAvg aggregation
```

```
    global_model = federated_average(
      global_model,
      local_models
    )
    print(
      " Global model updated "
      "using FedAvg"
    )
13. Evaluate Global Model
print("\nEvaluating Global EU-FMFI Model...\n")
test_data = generate_hospital_data(
  samples=600
)
accuracy, precision, recall, f1, cm = evaluate_model(
  global_model,
  test_data
)
print("---------------------------------")
print("EU-FMFI Performance")
print("---------------------------------")
print(
   f"Accuracy  : {accuracy:.4f}"
)
print(
  f"Precision : {precision:.4f}"
)
print(
 f"Recall    : {recall:.4f}"
)
print(
  f"F1-score  : {f1:.4f}"
)
print("\nConfusion Matrix:")
print(cm)
4. Uncertainty-Aware Prediction
print("\nUncertainty-Aware Prediction\n")
sample_image = test_data[0][0:1]
sample_ehr = test_data[1][0:1]
sample_text = test_data[2][0:1]
prediction, confidence, uncertainty, probability = \
 uncertainty_prediction(
global_model,
sample_image,
sample_ehr,
sample_text,
amples=20
  )
print(
  "Predicted class:",
  int(prediction[0])
)
print(
   "Confidence:",
   float(confidence[0])
)
print(
  "Uncertainty:",
  float(
 np.mean(uncertainty[0])
  )
)
print(
   "Class probabilities:",
   probability[0]
)
5. Clinical Decision Support
threshold = 0.70
if confidence[0] >= threshold:
  print(
 "\nClinical Decision:"
  )
  print(
    "Prediction has relatively "
    "high confidence."
  )
else:
   print(
     "\nClinical Decision:"
  )
   print(
     "Prediction has considerable "
     "uncertainty."
  )
   print(
      "Further clinical evaluation "
      "is recommended."
   )
print("\nEU-FMFI execution completed.")
```

5. Discussion

The results obtained for the proposed **Explainable and Uncertainty-Aware Federated Multimodal Foundation Intelligence (EU-FMFI)** framework provide an overall view of its learning behaviour, diagnostic performance, computational requirements, and potential usefulness in medical decision support. The analysis considers several measures rather than depending on accuracy alone. These include training and validation accuracy, loss, computational time, AUROC, ROC characteristics, the confusion matrix, precision, recall, and F1-score. This combination is important because medical diagnosis involves more than simply obtaining a correct classification. A useful diagnostic system should also be able to work with different forms of patient information, protect sensitive medical records, provide dependable predictions, and give clinicians sufficient information to judge whether a result is reasonable.

The framework is intended to bring together information obtained from medical images, electronic health records, laboratory findings, and clinical notes. These sources describe different aspects of a patient's condition and may therefore provide complementary evidence during diagnosis. The results presented in this study should nevertheless be interpreted with some care. The figures supplied for analysis contain illustrative numerical values, and there is a difference between the accuracy reported by the confusion matrix and the accuracy shown in the overall performance graph. Therefore, the numerical observations below describe the patterns represented in the supplied figures and should not be regarded as independently validated clinical results. For the final version of the research paper, all performance measures should be calculated from the same verified test dataset and experimental procedure.

5.1 Learning Behaviour of the Proposed Framework

The training results show a steady improvement in classification accuracy during the twenty epochs. Training accuracy starts at approximately 0.75 in the first epoch and gradually increases to about 0.965 by the twentieth epoch. Validation accuracy follows a similar pattern, beginning at approximately 0.72 and reaching around 0.91 by the end of training. The validation accuracy appears to reach its highest point, approximately 0.92, around the fifteenth epoch and then decreases slightly. This behaviour indicates that the model learns rapidly during the early stages of training, while the improvement on previously unseen validation data becomes smaller during the later epochs. The difference between training and validation accuracy is therefore useful when deciding when further training is no longer providing a meaningful improvement.

The loss curves support the same observation. Training loss decreases from approximately 0.83 in the first epoch to about 0.10 at epoch 20, while validation loss falls from approximately 0.95 to around 0.18. The continuous decline in both curves indicates that the model is progressively reducing its prediction error during training. At the same time, validation loss remains higher than training loss throughout the experiment, which is expected because the model is optimized using the training data. The relatively small increase in the gap between the two curves during the later epochs suggests that the model should not simply be trained for as many epochs as possible. In a properly controlled experiment, the validation results can be used to select a suitable stopping point and avoid unnecessary training.

5.2 Contribution of Multimodal Medical Information

The main strength of the proposed framework is its ability to consider different forms of medical information together. Medical images provide visual information about anatomical or pathological changes, whereas EHRs, laboratory results, medical histories, and clinical notes provide additional information about the patient's symptoms and clinical background. These sources may complement one another, particularly in cases where an image alone does not provide sufficient evidence for a reliable diagnosis. For example, two conditions may show similar visual characteristics but differ in their associated clinical history or laboratory findings. Considering these sources together can therefore provide a more complete basis for diagnostic assessment.

The AUROC comparison shown in the supplied results illustrates this progression. The reported AUROC increases from **0.88 for the single-modality approach** to **0.91 for centralized multimodal learning**, followed by **0.93 for federated multimodal learning** and **0.96 for EU-FMFI**. The trend suggests that incorporating information from multiple sources can improve the ability to distinguish between diagnostic categories. However, these values are part of the supplied illustrative results and should not be presented as experimentally validated performance without supporting data. The important observation is that multimodal diagnosis has the potential to provide richer clinical evidence, provided that information from different sources is integrated carefully and differences in data quality, availability, and clinical relevance are properly considered.

5.3 Federated Learning and Protection of Patient Information

Patient privacy is particularly important in medical research because healthcare records contain highly sensitive personal and clinical information. The proposed federated approach allows participating healthcare institutions to retain their patient data within their own environments rather than transferring complete records or medical images to a central repository. Hospitals can therefore participate in collaborative model development while maintaining greater control over the information collected from their patients. This approach is particularly relevant when several healthcare institutions wish to contribute to a common diagnostic system but cannot freely exchange patient-level information.

The computational-time results provide an additional view of how the proposed federated arrangement

behaves as the number of participating sites increases. According to the supplied graph, centralized training takes approximately **18 minutes for one site** and increases to approximately **270 minutes for 16 sites**. In comparison, the displayed federated EU-FMFI values increase from approximately **24 minutes to 118 minutes** over the same range. The graph therefore shows a slower increase in training time for the federated approach. However, computational time in a real healthcare environment depends on several factors, including the available hardware, network conditions, amount of local data, communication frequency, model size, and number of training rounds. Consequently, these values should be regarded as an illustrative comparison rather than sufficient evidence for making a general claim about the computational superiority of federated learning.

5.4 Role of Uncertainty in Medical Diagnosis

Medical diagnosis is rarely equally certain for every patient. Some cases may contain incomplete clinical records, poor-quality images, conflicting findings, or symptoms that overlap with several diseases. A system that simply produces one diagnostic label without indicating the level of confidence may therefore provide an incomplete picture. The proposed framework addresses this issue by considering uncertainty along with the predicted diagnosis. The purpose is to distinguish cases in which the available evidence strongly supports a particular diagnosis from cases where the evidence is less clear.

This distinction can be useful in clinical practice. A prediction supported by strong evidence and high confidence may provide useful additional support to a physician, whereas a prediction associated with considerable uncertainty may indicate that further tests or specialist assessment are necessary. For this reason, the proposed evaluation considers confidence and calibration alongside conventional diagnostic measures such as accuracy, precision, recall, F1-score, sensitivity, specificity, and AUROC. However, the supplied figures do not contain numerical uncertainty or calibration results. Therefore, it would not be appropriate to make a quantitative claim about the quality of the uncertainty estimates based on the current figures alone.

5.5 Explainability and Clinical Confidence

A diagnostic prediction is more useful to a clinician when there is some indication of the evidence behind it. In medical imaging, this may involve identifying the regions of an image that contributed most strongly to the prediction. For structured clinical information, relevant symptoms, laboratory findings, medical history, or other patient characteristics may provide the supporting evidence. Such information can help a doctor compare the system's prediction with the actual clinical presentation of the patient. The proposed framework therefore places importance on making the diagnostic process easier to examine rather than presenting only a final classification.

Explanations may also be helpful when the system produces an unexpected or incorrect result. If a model repeatedly bases its prediction on an irrelevant image region or an inappropriate clinical factor, this can provide an indication that the model requires further investigation. Combining an explanation with uncertainty gives clinicians two useful pieces of information: the explanation indicates **what evidence influenced the prediction**, while the uncertainty indicates **how strongly the available evidence supports that prediction**. This combination can make the output more informative and may help clinicians decide which cases can be considered relatively reliable and which require closer examination. The proposed research therefore considers relevance, consistency, and clinical usefulness when discussing explanation quality.

5.6 Class-Wise Classification Performance

The confusion matrix provides a more detailed picture of the classification results by showing where correct and incorrect predictions occur. In the supplied matrix, the diagonal values are **184 for Class 1, 171 for Class 2, and 181 for Class 3**. These values indicate that most of the cases shown in the matrix are assigned to their correct classes. The off-diagonal values provide additional information about the types of errors. In particular, **17 Class 2 cases are predicted as Class 3**, while **14 Class 3 cases are predicted as Class 2**. This indicates that Classes 2 and 3 have greater overlap in the supplied example than the other class combinations.

The class-wise results demonstrate why accuracy alone is not sufficient for evaluating a medical diagnostic system. Overall accuracy gives a general indication of performance, whereas precision and recall provide more information about how individual classes are being identified. The supplied overall performance graph reports an accuracy of **0.950**, precision of **0.940**, recall of **0.930**, and F1-score of **0.935**. These values indicate relatively consistent performance across the reported measures. However, they do not agree with the accuracy calculated from the supplied confusion matrix. Therefore, these results should not be presented as measurements from one common test set until the

underlying predictions have been verified and the graphs have been regenerated consistently.

5.7 Relevance to Clinical Decision Support

The significance of the proposed framework extends beyond obtaining a high classification score. In a real healthcare environment, a diagnostic system must deal with different forms of information, protect patient confidentiality, provide dependable results, and present those results in a form that healthcare professionals can understand. EU-FMFI attempts to address these requirements within a common framework by combining multimodal information, federated learning, uncertainty assessment, and explanation. This makes the proposed approach broader than a conventional classification model that provides only a predicted disease category.

The need for such an approach becomes more evident when different hospitals have different patient populations, medical equipment, record-keeping practices, and data quality. A model that performs well in one hospital may not necessarily behave in the same way in another setting. For this reason, the proposed evaluation considers robustness across participating institutions and communication requirements in addition to conventional diagnostic measures. However, actual evidence of generalization across hospitals would require experiments using real multi-institutional data. The present results therefore provide a basis for discussing this potential but do not by themselves establish general clinical generalizability.

5.8 Limitations of the Present Results

One of the main limitations of the current results is that the numerical values shown in the supplied figures are illustrative and are not accompanied by a documented clinical dataset, experimental protocol, or original prediction records. This issue is particularly important because the confusion matrix and the overall performance graph give different accuracy values. The confusion matrix contains **536 correctly classified cases out of 600**, which corresponds to approximately **89.3% accuracy**, whereas the overall performance graph reports **95.0% accuracy**. Both values cannot describe the same 600-case test set.

This inconsistency should be resolved before the results are included in a final journal manuscript. Ideally, the same verified set of predictions should be used to generate the confusion matrix, accuracy, precision, recall, F1-score, ROC curve, and AUROC. This would ensure that all reported values are directly comparable and traceable to the same experimental conditions. Such consistency is particularly important when presenting research for peer review, because readers should be able to understand exactly how each reported performance measure was obtained.

5.9 Future Research Directions

Future work should be based on clearly identified and properly documented medical datasets. The study should specify the patient population, data sources, inclusion and exclusion criteria, training and testing strategy, number of participating institutions or experimental sites, and the complete model-training procedure. Particular attention should be given to preventing information from the same patient from appearing in both training and testing groups. Such precautions are essential for obtaining a reliable estimate of diagnostic performance.

Further experiments should also include actual measurements of uncertainty, calibration, explanation quality, privacy protection, communication overhead, and performance across different healthcare institutions. These measurements would provide a more complete assessment of the framework than diagnostic accuracy alone. They would also help determine whether the proposed approach remains reliable when medical information is incomplete or when the characteristics of the data vary between hospitals. The research material identifies these aspects as important elements of the overall evaluation.

5.10 Overall Discussion

Taken together, the supplied results show the intended behaviour of the EU-FMFI framework across several aspects of medical diagnosis. The training curves indicate progressive improvement in accuracy, while the loss curves show a corresponding reduction in prediction error. The computational-time comparison illustrates how the displayed training requirements change as the number of participating sites increases. The AUROC and ROC results provide information about the ability to distinguish between diagnostic categories, while the confusion matrix gives a clearer picture of class-specific errors. Considering all of these measures together provides a more informative assessment than relying on a single performance indicator.The main contribution of the proposed framework is its attempt to bring **multimodal diagnosis, privacy-preserving collaboration, uncertainty assessment, and understandable diagnostic evidence** into one medical decision-support approach. Its purpose should be to assist healthcare professionals by providing additional information rather than replacing their clinical judgment. The practical value of the framework will ultimately depend on careful validation using real medical

datasets, consistent experimental procedures, appropriate privacy measures, and assessment by clinical users. With such validation, the proposed approach could provide a useful direction for developing medical diagnostic systems that are not only accurate, but also transparent, privacy-conscious, and suitable for responsible clinical use.

References

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. 20th Int. Conf. Artificial Intelligence and Statistics (AISTATS), vol. 54, 2017, pp. 1273–1282. https://proceedings.mlr.press/v54/mcmahan17a.html

[2] P. Kairouz et al., "Advances and open problems in federated learning," Foundations and Trends in Machine Learning, vol. 14, nos. 1–2, pp. 1–210, 2021. https://doi.org/10.1561/2200000083

[3] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," ACM Trans. Intelligent Systems and Technology, vol. 10, no. 2, pp. 1–19, 2019. https://doi.org/10.1145/3298981

[4] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," IEEE Signal Processing Magazine, vol. 37, no. 3, pp. 50–60, May 2020. DOI: https://doi.org/10.1109/MSP.2020.2975749

[5] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in Proc. Machine Learning and Systems (MLSys), vol. 2, 2020, pp. 429–450. https://proceedings.mlsys.org/paper_files/paper/2020/hash/1f5fe83998a09396ebe6477d9475ba0c-Abstract.html

[6] K. A. Bonawitz et al., "Practical secure aggregation for privacy-preserving machine learning," in Proc. ACM SIGSAC Conf. Computer and Communications Security, 2017, pp. 1175–1191. https://doi.org/10.1145/3133956.3133982

[7] R. C. Geyer, T. Klein, and M. Nabi, "Differentially private federated learning: A client level perspective," arXiv:1712.07557, 2017. https://arxiv.org/abs/1712.07557

[8] I. Dayan et al., "Federated learning for predicting clinical outcomes in patients with COVID-19," Nature Medicine, vol. 27, pp. 1735–1743, 2021. https://doi.org/10.1038/s41591-021-01506-3

[9] M. J. Sheller et al., "Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data," Scientific Reports, vol. 10, Art. no. 12598, 2020. https://doi.org/10.1038/s41598-020-69250-1

[10] K. D. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," Nature Machine Intelligence, vol. 2, pp. 305–311, 2020. https://doi.org/10.1038/s42256-020-0186-1

[11] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 2, pp. 423–443, Feb. 2019. https://doi.org/10.1109/TPAMI.2018.2798607

[12] Y.-H. H. Tsai et al., "Multimodal transformer for unaligned multimodal language sequences," in Proc. 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 6558–6569. https://doi.org/10.18653/v1/P19-1656

[13] J. Lu et al., "ViLBERT: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks," in Advances in Neural Information Processing Systems (NeurIPS), vol. 32, 2019. https://arxiv.org/abs/1908.02265

[14] A. Radford et al., "Learning transferable visual models from natural language supervision," in Proc. 38th Int. Conf. Machine Learning (ICML), vol. 139, 2021, pp. 8748–8763. https://proceedings.mlr.press/v139/radford21a.html

[15] J. Moor et al., "Foundation models for generalist medical artificial intelligence," Nature, vol. 616, pp. 259–265, 2023. https://doi.org/10.1038/s41586-023-05881-4

[16] K. Singhal et al., "Large language models encode clinical knowledge," Nature, vol. 620, pp. 172–180, 2023. https://doi.org/10.1038/s41586-023-06291-2

[17] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," in Proc. Int. Conf. Learning Representations (ICLR), 2021. https://openreview.net/forum?id=YicbFdNTTy

[18] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2022, pp. 16000–16009. https://doi.org/10.1109/CVPR52688.2022.01553

[19] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in Proc. 37th Int. Conf. Machine Learning (ICML), vol. 119, 2020, pp. 1597–1607. https://proceedings.mlr.press/v119/chen20j.html

[20] X. Chen and K. He, "Exploring simple siamese representation learning," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2021, pp. 15750–15758. https://doi.org/10.1109/CVPR46437.2021.01549.

[21] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 1135–1144. https://doi.org/10.1145/2939672.2939778

[22] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017. https://arxiv.org/abs/1705.07874

[23] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proc. IEEE Int. Conf. Computer Vision (ICCV), 2017, pp. 618–626. https://doi.org/10.1109/ICCV.2017.74

[24] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv:1702.08608, 2017. https://arxiv.org/abs/1702.08608

[25] A. Holzinger et al., "Causability and explainability of artificial intelligence in medicine," WIREs Data Mining and Knowledge Discovery, vol. 9, no. 4, 2019, Art. no. e1312. https://doi.org/10.1002/widm.1312

[26] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in Proc. 33rd Int. Conf. Machine Learning (ICML), vol. 48, 2016, pp. 1050–1059. https://proceedings.mlr.press/v48/gal16.html

[27] A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?" in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017. https://arxiv.org/abs/1703.04977

[28] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in Proc. 34th Int. Conf. Machine Learning (ICML), vol. 70, 2017, pp. 1321–1330. https://proceedings.mlr.press/v70/guo17a.html

[29] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017. https://arxiv.org/abs/1612.01474

[30] J. Lee et al., "BioBERT: A pre-trained biomedical language representation model for biomedical text mining," Bioinformatics, vol. 36, no. 4, pp. 1234–1240, 2020. https://doi.org/10.1093/bioinformatics/btz682