Proceedings of the 4th International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS-2026)
DVD Part Number: CFP26DN7-DVD; ISBN: 979-8-3315-7008-8

Causal-Aware Federated Neuro-Semantic Capsule Transformer Framework for Robust and Early Prediction of Meningitis

A. Shabana
Research Scholar
Department of Computer Science,
School of Computing Sciences
Vels Institute of Science, Technology
& Advanced Studies (VISTAS)
Chennai, India
shabanaabbas1099@gmail.com

P. Kavitha
Assistant Professor
Department of Computer Applications
School of Computing Sciences
Vels Institute of Science, Technology
& Advanced Studies (VISTAS)
Chennai, India
pkavikamal@gmail.com

***Abstract* —Meningitis is a severe neuroinfectious disease that can cause serious neurocognitive impairment and incurable rates of death when it is not diagnosed early. Existing diagnostic and forecasting techniques rely on manual clinical examination and laboratory analyses, which are time-intensive, error-prone, and, in most cases, yield high rates of false positives. To circumvent these shortcomings, this paper proposes next-generation deep learning and optimisation approaches within a higher-level artificial intelligence framework to predict meningitis with high accuracy and at an early stage. The proposed model incorporates a Neuro-Semantic Capsule Transformer Network (NSCTN), which uses a combination of a capsule-based spatial hierarchy and transformer-based contextual learning to learn complex interactions among clinical, biochemical, and neuroimaging features. Before classification, an Adaptive Distribution Harmonisation (ADH) strategy is used to normalise the data, aiming to remove scale imbalance and noise. The Deep Causal Impact Scoring (DCIS) mechanism is used to assess feature relevance, as it measures causal impact rather than simple correlation. To further optimise feature selection and convergence efficiency, an Adaptive Bayesian Hypergraph Attention Optimiser (ABHAO) is used, which allows modelling of multi-feature interactions while preserving probabilistic uncertainty. Moreover, a Causal-Aware Federated Meta Learning Network (CAFMLN) is proposed to enhance generalisation and the robustness of heterogeneous clinical datasets without compromising data privacy. Through experimental confirmation, the proposed framework is shown to be far better than traditional machine learning, deep neural networks, and state-of-the-art models in terms of accuracy, precision, recall, F1-score, and false alarm rate. The model also has better generalisation properties and robustness to different data distributions.**

Keywords— Neuro-Semantic Capsule Transformer Network (NSCTN), Adaptive Distribution Harmonisation (ADH), Deep Causal Impact Scoring (DCIS), Adaptive Bayesian Hypergraph Attention Optimiser (ABHAO), Causal-Aware Federated Meta Learning Network (CAFMLN)

I. INTRODUCTION

Forecasting complex events quickly and accurately is challenging in healthcare and other fields that are constantly evolving. When using traditional methods to predict future events, predictive value and accuracy typically suffer from the use of high-dimensional, non-linear, and time-varying data, leading to higher error rates. Recent advances in deep generative modelling techniques, such as Generative Adversarial Networks (GANs), have shown great promise in producing improved performance in the areas of feature representation and approximate synthetic data generation, as well as uncertainty quantification for applications such as energy systems, network performance analysis, and clinical patient monitoring through capturing the complex dynamics of temporal trends and multimodal interactions. Due to their ability to capture the complex interactions in time-varying data and the large diversity of uncertainties across different dimensions, these models are highly appropriate for medical predictions. Meningitis is a severe neuroinfectious disease that is associated with high mortality and severe complications within the nervous system. Rapid diagnosis of meningitis is critical to improve patient outcomes. Conventional diagnostic methods for meningitis rely on laboratory testing and clinical judgement, which frequently lead to delayed decisions about how to treat a patient (e.g., misdiagnosis) and false negatives resulting from either laboratory or clinical errors. To address the issues associated with conventional diagnostic methods, this study introduces an advanced artificial intelligence (AI) framework for predicting and classifying meningitis at its early stages. Due to delayed diagnosis and limited access to expert systems, early prediction of meningitis remains a significant concern. Current centralised models suffer from poor cross-institutional generalisation and data privacy issues. A solid, privacy-preserving, and causally based prediction framework is thus required.

A Causal-Aware Federated Neuro-Semantic Capsule Transformer (CAFNSCT) method has been developed for early diagnosis of meningitis. ADH, DCIS, ABHAO, NSCTN, and CAFMLN combination for enhanced feature extraction and causality-based learning. The architecture of the AI framework utilizes a Neural-Semantic Capsule Transformer Network (NSCTN) for modelling the hierarchical and contextual clinical relationships associated with patients with meningitis, a Deep Causal Impact Scoring Mechanism (DCIS) for providing causality information for contributing features, and an Adaptive Bayesian Hypergraph Attention Optimizer (ABHAO) for capturing the interactions of the contributing features across the patients involved in the study. Furthermore, a Causal-Aware Federated Meta Learning Network (CAFMLN) is introduced into the overall

architecture to improve the generalisability of predictions across disparate datasets, while preserving the confidentiality and privacy of patients' information. The focus of the overall system is to decrease false-positives, to improve predictive accuracy, and to provide a greater interpretability of patients' conditions through the use of causal analysis. Experimental evaluations show that the overall system out-performs traditional machine learning and deep learning methods with respect to precision, recall, F1-score, and robustness.

II. RELATED WORK

Recent research indicates that the influence of advanced deep learning models has been increasing in various fields of predictive and diagnostic models. Berghout et al. (2023) came up with a collaborative augmented hidden layer feedforward neural network (RCFE-NAHL) to lithium-ion battery State of Health prediction. They combine a Robust Collaborative Feature Extractor with augmented Restricted Boltzmann Machines to learn intricate ageing behaviour, and augment this with an augmented hidden layer network to learn nonlinear degradation behaviour to provide better predictive accuracy. Song et al. (2018) proposed Multimodal Stochastic Recurrent Neural Networks (MSRNNs) to video captioning in video analytics. Their generative model one approach to uncertainty in a temporal and semantic context by using stochastic latent variables and multimodal inputs can generate a wider range of captions and provides a richer context of captions than deterministic RNNs. Likewise, Amanullah et al. (2025) have created a hybrid meningitis diagnosis model based on the integration of Fuzzy Cognitive Maps with Decision Trees, which has improved the interpretability and the misclassification rate of simulated and real clinical data.

There is also improvement in robustness and planning applications. Doan et al. (2022) tested universal naturalistic adversarial patches (TNT attacks) that caused major inaccuracies in neural networks and highlighted the vulnerabilities to security. Zhao et al. (2022) have used an ensemble of Generative Adversarial Networks and Rapidly-Exploring random Trees to plan efficient and safe trajectories. Moreover, Franchi et al. (2023) enhanced uncertainty quantification in Bayesian Neural Network through latent posterior encoding, and Bessadok et al. (2022) emphasized the usefulness of Graph Neural Network in the neurological disease modelling. Also, a CNN-based sleep stage classifier that uses photoplethysmography signals was developed by Habib et al. (2022) and delivered a performance equivalent to that of clinical polysomnography.

TABLE I. COMPARATIVE ANALYSIS OF RECENT ADVANCES IN GENERATIVE AND DEEP LEARNING APPLICATIONS ACROSS DIVERSE DOMAINS

Ref.	Domain	Methods	Data Type	Key Findings
Striuk et al., (2021)	Generative models survey	GANs and deep learning	Data Authentication	Summarises successful GAN cases & advanced architectures.
Vital Jr et al. (2025)	Citation impact prediction	GPT & text embeddings	Scientific papers meta-data	Text embeddings effectively predict citation impact.
Ivanovic et al. (2020)	Trajectory prediction	Conditional Variational Autoencoder (CVAE)	Multimodal trajectory data	CVAE captures uncertainty and multimodality..
Chava & Saradhi (2024)	Pharmaceutical supply chains	Generative AI & deep neural networks	Supply chain datasets	Shows generative AI enables automated insights
Tao et al. (2025)	Battery capacity estimation	Deep generative transfer learning	Second-life Li-ion batteries	Transfer learning enhances estimation accuracy across heterogeneous battery sources
Errezgouny et al. (2025)	Predictive vehicle maintenance	Integrated deep learning framework	Vehicle sensor/ maintenance data	Improves maintenance prediction and reduces unplanned downtime
Hussain et al. (2025)	Industrial manufacturing	Deep learning & generative AI	Manufacturing workflow & control data	AI accelerates prototyping, improves export control strategies, and strengthens IP
Topşir et al. (2025)	Medical classification	GANs + Kolmogorov-Arnold Network	Thyroid disease clinical records	Achieves high three-class classification accuracy with synthetic data augmentation
Becker et al. (2025)	Condition monitoring	Explainable neural networks	Time-series from sheet metal shearing	Provides interpretable monitoring with strong predictive performance

Table 1 provides an overview of the recent work in the application of generative and deep learning models in a variety of fields, such as scientific prediction, trajectory modelling, battery estimation, healthcare, manufacturing, and supply chains. The accuracy, decision-making, efficiency, and interpretability have been improved in these works, and they emphasize the generative adversarial and explainable neural networks versatility in the complex real-life applications.

III. PROPOSED METHODOLOGY

The proposed methodology presents a novel AI architecture for early meningitis prediction that leverages deep learning, causal inference, and optimisation algorithms. Adaptive Distribution Harmonisation is then used to refine clinical, biochemical, and neuroimaging data to ensure quality and consistency. Deep Causal Impact Scoring identifies the truly influential features, and the Adaptive Bayesian Hypergraph Attention Optimiser selects informative attributes by modelling high-order interactions and uncertainty. The proposed framework operates in four

stages: (i) preprocessing and normalisation, (ii) feature extraction using DCIS, (iii) causal representation learning via NSCTN, and (iv) federated optimisation using CAFML. These attributes are trained using a Neuro-Semantic Capsule Transformer Network to gain hierarchical and global relationships. Lastly, a Causal-Aware Federated Meta Learning Network helps train across various datasets, achieving high accuracy, robustness, and fewer false alarms.

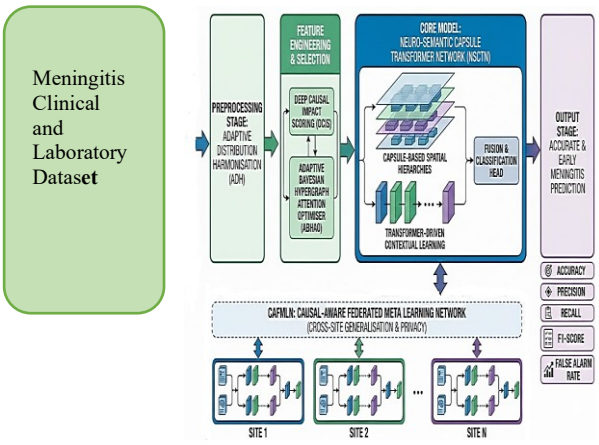

Fig. 1. Proposed AI-Driven Workflow for Early Meningitis Prediction Using NSCTN and CAFMLN

The proposed meningitis forecasting framework is described in Figure 1. Adaptive Distribution Harmonisation The heterogeneous clinical data is preprocesses by Adaptive Distribution Harmonisation and then causally directed feature selection by Deep Causal Impact Scoring and the Adaptive Bayesian Hypergraph Attention Optimiser. The refined features are then subjected to Neuro-Semantic Capsule Transformer Network and Causal-Aware Federated Meta Learning is used to guarantee privacy-preserving and generalisable prediction.

A. Adaptive Distribution Harmonisation (ADH)

The Adaptive Distribution Harmonisation (ADH) method is used in preprocessing to guarantee the high quality and variability of the data. By capturing higher-order dependencies, which are frequently overlooked in traditional graph-based models, the DH module improves relational feature learning. To systematically handle missing and null values, the ADH is processed by estimating feature-wise statistical distributions, then filling incomplete entries with distribution-based approximations that preserve the intrinsic data properties. It also uses adaptive thresholding to detect noise and outliers, accounting for global and local feature variations, thereby achieving effective anomaly removal without interfering with meaningful patterns. Also, the feature heterogeneity is handled by ADH by dynamically normalizing the features into a common representational space. ADH increases the numerical stability and acceleration of convergence in the meningitis diagnosis process and increases the predictive reliability by reducing redundancy and balancing feature distributions.

$$x_{ij}^{*} = \begin{cases} x_{ij}, & \text{if } x_{ij} \neq \emptyset \\ \mu_j, & \text{if } x_{ij} = \emptyset \end{cases} \quad (1)$$

In Equation (1), x_{ij} rep denotes the initial value of the j th feature of the i th data sample in the raw data. $x_{ij}+$ is a symbol that is used to represent the revised or imputed value of a feature after preprocessing with Adaptive Distribution Harmonisation (ADH) method. The $\emptyset$ symbol is used to represent an absent or null record in the data set. The $\mu_j$represents the statistical mean of the j th feature, which is calculated using all non-missing samples. In case a feature value has not been observed, x_{ij}^{*}maintains the same value as x_{ij}.

$$\sigma_j^2 = \frac{1}{N}\sum_{i=1}^{N}(x_{ij} - \mu_j)^2 \quad (2)$$

In the Equation (2), σ_j^2 represents the variance on the j th feature, which is a measure of how far the values of a feature are distributed about the mean. Nrep is a variable that describes the number of data samples in the dataset. The x_{ij} represents the value of the j feature of the i sample whereas, μ_j represents the mean value of the j -th feature calculated over all samples. Summation $\sum_{i=1}^{N}(x_{ij} - \mu_j)^2$captures both global and local variability in the values of each feature: it is the squared difference between the value of an individual feature and its mean.

$$x_{ij}^{\text{clean}} = \begin{cases} x_{ij}, & |\, x_{ij} - \mu_j \,| \leq \lambda\sigma_j \\ \mu_j, & \text{otherwise} \end{cases} \quad (3)$$

Where (3) means, x_{ij} clean, means the value of the j feature of the i sample after eliminating the outliers. x_{ij} is the original feature value and σ_j are the mean and variance of the j th feature respectively. The λ is an adaptive scaling parameter which ascertains the allowable degree of variance with the mean. When the value of the feature is within λ times the standard deviation about the mean, it is kept otherwise the replaced with the mean is done to remove extreme outliers.

$$w_j = \frac{1}{1+\sum_{k\neq j}\text{Cov}(x_j, x_k)} \quad (4)$$

Equation (4) w_j denotes the redundancy weight of the j feature, and is applied to minimize correlated or redundant information. $cov(x_{j,x_k})$ is the covariance between the j feature and the other features $k \neq j$ The use of the sum of covariances in the denominator, makes those features that are highly correlated with the rest of the features have a lower weight which ensures that those features that are independent and informative are highlighted in preprocessing.

$$x_{ij}^{\text{norm}} = \frac{x_{ij}^{\text{clean}} - \mu_j}{\sigma_j} \quad (5)$$

The meaning of this term is as follows: x_{ij}^{norm} is the normalised value of the j feature of the i sample x_{ij}^{clean} is the cleaned value of the feature of Equation (5), and the μ_j are the mean and standard deviation of the feature. This normalisation σ_j is a standardisation of the feature, which makes the feature have a zero mean and a unit variance, thus allowing the homogenisation of the heterogeneous features.

$$x_{ij}^{\text{ADH}} = \frac{x_{ij}^{\text{norm}} - \min(x_j)}{\max(x_j) - \min(x_j)} \quad (6)$$

In this case, x_{ij}^{ADH} is the final harmonised value of the j^{th}feature following Adaptive Distribution Harmonisation; x_{ij}^{norm} is the normalised feature value of Equation (6), and (x_j) are the minimum and maximum values of the feature, respectively. This minmax scaling is used to normalize all the values of features to the same range, usually [0,1] so that the values of different features can all be compared.

$$H(X) = -\sum_{j=1}^{M} p(x_j)\log\, p(x_j) \quad (7)$$

In the equation (7), H(X) rep is the amount of entropy of the dataset following the preprocessing, which measures the distributional uncertainty of features. M is the total number of features, and $p(x_j)$ is the probability distribution of the j^{th} feature values. Reduced entropy values reveal a more balanced and similar data set, which proves that ADH has been effective in removing noise, outliers and heterogeneity of features.

B. Deep Causal Impact Scoring (DCIS)

Deep Causal Impact Scoring (DCIS) is a quantitative measure of actual causal impact of every input feature on meningitis prediction after preprocessing. In comparison to traditional correlation-based selection technologies, which can be used to select spurious relationships, DCIS directly models cause effect relationships between clinical variables and outcomes. It approximates the direct impact of the variation of features by merging deep neural representations with causal principles of inference and relying on confounders. DCIS uses interventional simulations and counterfactual analysis to give impact scores, which represent true predictive value. Powerful causal characteristics are promoted, and others are minimised or altogether suppressed, promoting interpretability, bias reduction, and robustness and generalisation.

$$\mathbf{z}_j = f_\theta(x_j) \quad (8)$$

In Equation (8), z jrep is the deep feature embedding of the j^{th} input feature transformed by the neural network. The parameterisation $f_\theta(\cdot)$represents a learnable deep neural mapping that represents intricate nonlinear correlations between input features. The x_j is the preprocessed value of the j^{th} feature which is received at the Adaptive Distribution Harmonisation (ADH) step.

$$P(y \mid do(x_j)) = \sum_{\mathbf{z}_j} P(y \mid \mathbf{z}_j)\, P(\mathbf{z}_j \mid do(x_j)) \quad (9)$$

In Equation (9), $P(y \mid do(x_j))$ is the interventional probability of the target event, y (meningitis or non-meningitis) when the j th feature, j^{th}, is set or intervened upon externally. It is summed across all potential deep features embeddings z_j, that are learnt representations of x_j obtained via neural mapping $f_\theta(x_j)$. $P(y \mid z_j)$ the conditional probability of outcome y given feature embedding z_j, while $P(z_j \mid do(x_j))$is the probability of feature embedding z 1given the intervention $do(x_j)$.

$$P(y) = \mathbb{E}[P(y \mid X)] \quad (10)$$

$P(y)$ is used in Equation (10) to denote the baseline or marginal probability of the target outcome y (meningitis or non-meningitis) in the entire dataset. $E[P(y \mid X)]$ is used to indicate the expected value of the corresponding conditional probability of y when all input features are $X = \{x_1, x_2, \dots, x_M\}$, and so on, is the entire repertoire of preprocessed features of any given sample of data. The model allows quantifying how sincere causal input of individual features influences the prediction of meningitis by comparing $P(y \mid do(x_j))$ and $P(y)$.

$$\Delta_j = \mid P(y \mid do(x_j)) - P(y) \mid \quad (11)$$

In equation (11) Delta jrep is the marginal causal effect of j^{th} feature x_j on the target set y (meningitis or non-meningitis). It is computed by the absolute difference between the interventional probability, $P(y \mid do(x_j))$, the probability of outcome given the externality of x_j and the baseline probability, $P(y)$. This difference is a quantification of the direct effect of feature x_j the prediction which removes the effects of confounding variables or indirect associations.

$$\omega_j = \frac{1}{1+\sum_{k\neq j}\text{Conf}(x_j,x_k)} \quad (12)$$

In Equation (12) 0 jrep signifies the causal weight of adjustment of the j^{th} feature x j, it is utilized to determine the effect of confounding factors when estimating the causal impact. $\text{Conf}(x_j, x_k)$ is used to measure the level of confounding or dependency between feature x_j and all other features x_k in the dataset, where $k \neq j$. The aggregation of the confounding effect of all the other features is $\sum_{k\neq j}\text{Conf}(x_j, x_k)$ The model can be used by computing $\omega_j = 1/(1 + \sum_{k\neq j}\text{Conf}(x_j, x_k))$ Conf and then the down weight features that are strongly confounded by others, so that the estimated causal effect can be on the actual direct effect of x_j on the outcome y. This modification increases the accuracy and interpretability of feature weights in the prediction of meningitis.

$$\text{DCIS}_j = \omega_j \cdot \Delta_j \quad (13)$$

The expression of the Deep Causal Impact Score of the j^{th} feature x_j, in Equation (13) as DCIS_j is the actual causal impact of the feature x_j, on the prediction of meningitis. The marginal causal effect of the feature is denoted by the terms Δ_j denotes the absolute change in probability of the outcome as a result of the feature being manipulated, x_j,relative to the probability without manipulation. Variable omega j is the adjustment weight that gives a 100 percent weight to the confounding effects of all other features, and it makes sure that the score is not a result of the spurious relationships.

C. Adaptive Bayesian Hypergraph Attention Optimiser (ABHAO)

After conducting the causal analysis Adaptive Bayesian Hypergraph Attention Optimiser (ABHAO) is used to identify most informative and complementary features in meningitis prediction. ABHAO is a multidimensional relationship model between interacting clinical variables based on hypergraph attention, unlike the conventional method which analyses features separately. This method assigns the weights of importance dynamically based on the contribution and the strength of interaction, and, therefore, critical latent dependencies are discovered. At the same time, Bayesian optimisation is

used to direct feature selection in case of uncertainty by revising probabilistic beliefs of feature relevance. In terms of balanced exploration and exploitation, ABHAO eliminates overfitting and premature convergence, eventually producing an optimal subset of features which can be more stable, faster to converge and more accurate in prediction.

$$\mathrm{H}(v,e) = \begin{cases} 1, & \text{if } v \in e \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

The adaptive Bayesian hypergraph attention optimiser (ABHAO) has the feature nodes and hyperedges related by the matrix $H(v,e)$ that is shown in Equation (14). The v is the feature node representing the v^{th} th clinical attribute, and erep denotes a hyperedge, which relates to a group of mutually dependent features. The value of $H(v,e)$will be established to be 1 when feature v is contained in the hyperedge e, which implies a direct relationship, and 0 otherwise, which implies there is no association between the two.

$$\alpha_{ij} = \frac{\exp(\phi(x_i,x_j))}{\sum_{k\in e_j}\exp(\phi(x_k,x_j))} \quad (15)$$

In Equation (15), α_{ij} is the attention weight that the i^{th} feature node x_i will receive in the hyperedge e_j in the Adaptive Bayesian Hypergraph Attention Optimiser (ABHAO). Function $\phi(x_i,x_j)$ a learnable compatibility or scoring function, an estimate of the relevance or strength of interaction of feature x_i and feature x_j. These scores are normalised by the denominator $\sum_{k\in e_j}\exp(\phi(x_k,x_j))$ to condition that the attention scores add up to one, but only across feature nodes k within each hyperedge e_j,. Computation of α_{ij} dynamically measures the relative importance of each attribute in the scenario of its interaction with other attributes to allow selecting very informative and interdependent attributes in meningitis prediction.

$$P(x_j \mid \mathcal{D}) \propto P(\mathcal{D} \mid x_j)\,P(x_j) \quad (16)$$

In Equation (16) $P(x_j \mid \mathcal{D})$is the posterior probability that the j^{th}feature x_j is relevant in the given observed data set Din the Adaptive Bayesian Hypergraph Attention Optimiser (ABHAO). $P(\mathcal{D} \mid x_j)$is the probability used to indicate how well the observed data Dcan be explained by the value of the feature $x_j$and $P(x_j)$is the prior probability that is used to reflect any prior belief of how important the feature is before looking at the data. Proportionality The symbol $\propto$ is used to show that the posterior is calculated to a normalising constant. The given Bayesian formulation will enable ABHAO to revise beliefs of feature relevance in the face of uncertainty to perform uncertainty-aware feature selection and prioritise those attributes with high diagnostic power in predicting meningitis.

$$S_j = \alpha_j \cdot P(x_j \mid \mathcal{D}) \quad (17)$$

The adaptive selection score of the j^{th} feature x_j in Adaptive Bayesian Hypergraph Attention Optimiser (ABHAO) is represented by S j rep in Equation (17). The hypergraph attention weight of x_j denoted 0 j identifies the relative importance of x_j relative to its interactions with other features in the same hyperedge, and $P(x_j)$ is the posterior probability of the feature being relevant given the observed dataset D, computed by performing Bayesian inference. Through attention-based importance and uncertainty-aware probabilistic relevance, the selection score $S_j = \alpha_j \cdot P(x_j \mid \mathcal{D})$ helps to use a robust measure to select both highly informative and interdependently important features in predicting meningitis with high precision, allowing the optimiser to identify the most informative attributes to predict meningitis.

$$\mathcal{F}^* = \arg\max_{\mathcal{F}} \sum_{x_j\in\mathcal{F}} S_j \quad (18)$$

In Equation (18), $\mathcal{F}^*$represents an optimal feature set that the Adaptive Bayesian Hypergraph Attention Optimiser will choose in predicting meningitis in men. S_j is the adaptive selection score of the j^{th} feature that is the combination of the hypergraph attention weights and the Bayesian posterior probabilities to measure its significance. The summary $\sum_{x_j\in\mathcal{F}} S_j$is the sum of the cumulative score of all features in a candidate subset F. $ABHAO$ determines the set $\mathcal{F}^*$ 0 of features that maximises the total relevance and, in the process, includes high-order interactions between features, by optimising an argument of the maximum, $\mathcal{F}^*$max, over subsets of features.

D. Neuro-Semantic Capsule Transformer Network (NSCTN)

The features which are optimally chosen are then entered into Neuro-Semantic Capsule Transformer Network (NSCTN) to learn the deep representations and then do classification. NSCTN integrates network of capsules and transformer networks to identify local hierarchical and global contextual dependencies in data of meningitis. Capsule layers are used to retain patterns of spatial and relational features, and self-attention models based on transformers are used to transform long-range semantic interactions on clinical attributes. This hybrid design has produced strong noise resistant representations, improving the differentiation between meningitis and non-meningitis cases whilst having good generalisation across a variety of clinical data.

$$\mathrm{u}_i = g(\mathrm{W}_i\mathrm{x} + \mathrm{b}_i) \quad (19)$$

The first term in the Equation (19) is u_i which is the primary capsule output vector of the i^{th} feature of the Neuro-Semantic Capsule Transformer Network ($NSCTN$). Where x is the input feature vector (gained during the feature selection step), and W_i and $b_i$are the learnable weight matrix and the bias vector of the i^{th}capsule. The activation $g(\cdot)$ is a nonlinear activation considered to be ReLU or sigmoid, and it brings nonlinearity to the transformation. The network, by calculating $u_i = g(W_i x + b_i)$, transforms the input features to high-dimensional capsule representations which encode all the parameters of patterns at once, both presence and instantiation.

$$\mathrm{v}_j = \mathrm{squash}\Big(\sum_i c_{ij}\,\mathrm{u}_i\Big) \quad (20)$$

In Equation (20), $\mathrm{v}_j$is the output of the j th capsule following the dynamic routing in the the Neuro-Semantic Capsule Transformer Network ($NSCTN$). The $\mathrm{u}_i$the input vectors of the lower-level capsules and c_{ij} denote routing coefficients which express the value that each lower-level capsule u gives to the higher-level capsule v_j. These contributions are added together with

$\left(\sum_i c_{ij}\, \mathrm{u}_i\right)$ aggregates the weighted effect of all the lower-level features which are relevant. An example of a nonlinear vector normalisation is the squash function, which normalises the magnitude of v_j. between 0 and 1 without changing its orientation so that the output contains both the probability that the feature appears and its instantiation parameters.

$$\text{Attention}(\mathrm{Q},\mathrm{K},\mathrm{V}) = \text{softmax}\left(\frac{\mathrm{QK}^{\mathrm{T}}}{\sqrt{d_k}}\right)\mathrm{V} \quad (21)$$

In the Equation (21), Attention (Q, K, V) is the self-attention mechanism displayed in the transformer modules in Neuro-Semantic Capsule Transformer Network $(NSCTN)$. $Q(queries), K(keys)$, and $V(values)$ matrices are based on the input features representations and they store data on the relationships of the features. The dimensionality of the key vectors is denoted by d_k and used to scale the dot product $\mathrm{QK}^{\mathrm{T}}$to ensure that large values do not occur, which would lead to the destabilisation of the softmax calculation. The scaled dot product is normalised by the softmax function, which results in attention weights, which determine the extent to which a certain feature should pay attention to other features in the sequence.

$$\hat{y} = \mathrm{z}\sigma(\mathrm{W}_o + \mathrm{b}_o) \quad (22)$$

The prediction of the Neuro-Semantic Capsule Transformer Network $(NSCTN)$ output is denoted $\hat{y}$ in the Equation (22), where y widehat denotes the probability of a sample to fall under meningitis or non-meningitis. The variable z is the representation of integrated feature, which is a result of the summation between capsule outputs and transformer-based embeddings. The learnable output layer weights and bias of the output (high-dimensional feature z) to the prediction space are the matrix W_o and the vector b_o. The $\mathrm{z}\sigma(\mathrm{W}_o + \mathrm{b}_o)$operation is an activation operation, e.g. sigmoid to classify using two categories, or softmax to classify using many categories.

E. Causal-Aware Federated Meta Learning Network (CAFMLN)

Causal-Aware Federated Meta Learning Network (CAFMLN) is a network that improves generalisation and at the same time maintains high patient privacy. It allows distributed training between distributed clinical institutions, in which local models are trained on different systems, without the exchange of explicit patient data. Model updates and causal representations are only sent to a central coordinator and it meets the requirements of healthcare privacy. CAFMLN emphasizes meaningful patterns, and not spurious correlations by creating causal awareness into federated learning. It has a meta-learning aspect, which enables quick adaptation to new clinical settings, enhances transferability, minimizes overfitting, and preserves the diagnostic performance of diverse patient groups.

$$\theta_k^{(t+1)} = \theta^{(t)} - \eta\nabla_\theta\mathcal{L}(\theta^{(t)};\mathcal{D}_k) \quad (23)$$

In Equation (23), theta $\theta_k^{(t+1)}$is the revised model parameters of the K^{th}local client in the Causal-Aware Federated Meta Learning Network $(CAFMLN)$ at time $(t+1)$. The global model parameters that are shared with the client at iteration t are denoted by 0, and $\mathcal{D}_k$ is the local dataset of the k th institution. $\mathcal{L}(\theta^{(t)};\mathcal{D}_k)$ Is the loss function calculated on the local data, and 0 theta $\mathcal{L}(\theta^{(t)};\mathcal{D}_k$is the gradient of the loss on the model parameters. The learning rate η is what determines the size of the step.

$$\psi_j = P(y \mid do(x_j)) \quad (24)$$

In formula (24) ψ_j is the score of causal relevance of thej^{th} th feature x_j in Causal-Aware Federated Meta Learning Network $(CAFMLN)$. It is characterized as the interventional probability of $P(y \mid do(x_j))$, and is expressed as the likelihood of getting that outcome when x is set or manipulated externally. The model uses the evaluation of ψ_j to determine the extent to which each feature affects the prediction so as to isolate the direct causal impact but hold constant confounding variables.

$$\theta^{(t+1)} = \sum_{k=1}^{K} \frac{|\mathcal{D}_k|}{\sum_{k=1}^{K}|\mathcal{D}_k|}\,\theta_k^{(t+1)} \quad (25)$$

In (25), the result of the previous iteration t is represented by $\theta^{(t+1)}$in CAFMLN$\theta_k^{(t+1)}$is the locally updated parameters obtained from the k-th client's local dataset, where D_k represents the size of the local dataset. The weight $\frac{|D_k|}{\sum_{k=1}^{K}|D_k|}$ allows clients who have larger datasets to contribute more to the global update than those who have smaller datasets.

$$\theta^* = \arg\min_\theta \sum_{k=1}^{K}\mathcal{L}\left(\theta - \alpha\nabla_\theta\mathcal{L}(\theta;\mathcal{D}_k)\right) \quad (26)$$

In Equation (26), $\theta^*$the optimal parametric state of the Meta-Learnt model of the CAFMLN. The form of $L(\theta - \alpha\nabla_\theta L(\theta; D_k))$ is the evaluated loss function on the k^{th}client's Local dataset D_k after performing an approximate gradient descent based on the specified Meta learning rate α. $\sum_{k=1}^{K}$ the Loss for all K Federated Learning Clients. The optimal Updating Parametric State θ^*minimises the aggregated loss for all participating clients. Thus, allowing the overall model to adapt quickly to New or Unknown datasets.

IV. RESULT AND DISCUSSION

The experimental findings show that the offered AI-based meningitis prediction model achieves a higher success rate than traditional machine learning, conventional deep neural networks, and the latest models. The Neuro-Semantic Capsule Transformer Network, fuelled by Adaptive Distribution Harmonisation, learns more effectively from heterogeneous clinical, biochemical, and neuroimaging data, achieving higher accuracy, precision, recall, and F1-score, and fewer false alarms. Adaptive Bayesian Hypergraph Attention Optimiser and Deep Causal Impact Scoring provide additional information on feature relevance and convergence, as they model causal and multidimensional interactions. Also, the Causal-Aware Federated Meta Learning Network achieves strong generalisation across a wide array of datasets while maintaining patient privacy, facilitating practical clinical application. The recommended structure improves the accuracy of typical deep learning techniques while minimising computational overhead, making it ideal for environments with limited resources.

The proposed federated learning scheme can be implemented across different hospitals and institutions, where data remain at the hospital level while model parameters are shared. This method enables collaboration without compromising data confidentiality. It also reduces the amount of data communicated, enabling scalability and efficient processing across different hospitals. It involves distributed training across multiple clients rather than loading all data onto a single server.

A. Dataset Description: Meningitis Clinical and Laboratory Dataset

The Meningitis Clinical and Laboratory Dataset is a set of patient-level clinical and laboratory data, including demographics, vital signs, biochemical markers, haematological parameters, and other pertinent medical observations. These inhomogeneous features facilitate medical diagnosis, risk assessment and optimal prognosis. The dataset is rich and diverse, well-suited for machine learning and deep learning classification tasks to develop accurate diagnostic models and support clinical decision-making in the management of meningitis. This database consists of two classes. One is the Meningitis class, which denotes positive samples; the other is the non-meningitis class, which denotes negative samples. There are several samples in the positive class and B samples in the negative class, resulting in an [un]balanced distribution. Several preprocessing steps were implemented before training to improve model performance. Missing and corrupting samples in the dataset were first eliminated. Input data normalisation was performed using values ranging from 0 to 1.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Patient_ID	Age	Gender	WBC_Cou	Protein_Li	Glucose_L	Pathogen	Diagnosis	Outcome	Hemoglol	WBC_Bloc	Platelets
2	1	66	Female	14450	179	19	Yes	Bacterial	Recovere	3	13912	119405
3	2	94	Male	13470	122	104	No	Bacterial	Recovere	16	6845	213495
4	3	23	Male	8921	20	66	No	Viral	Recovere	18	4049	217301
5	4	53	Female	16200	145	16	Yes	Bacterial	Recovere	1	17731	119570
6	5	47	Female	4781	43	58	No	Viral	Recovere	14	5086	259521
7	6	75	Female	14127	107	1	Yes	Bacterial	Recovere	1	12228	118669
8	7	74	Female	13162	174	24	Yes	Bacterial	Recovere	6	13753	121651
9	8	92	Male	2354	163	4	Yes	Viral	Deceased	6	16392	142968
10	9	47	Female	15328	170	26	Yes	Bacterial	Recovere	8	17408	140540

Fig. 2. Sample Representation of the Meningitis Clinical and Laboratory Dataset

Figure 2 demonstrates the sample of the Meningitis Clinical and Laboratory Dataset, which includes patient level data including demographics, laboratory parameters, presence of the pathogen, type of diagnosis, and outcomes. Having both numeric and nominal characteristics, the heterogeneous data enables the use of the supervised learning in predicting the type of meningitis, its severity, and clinical outcomes with the help of the proposed AI framework

Comparison results

TABLE II. PERFORMANCE COMPARISON OF EXISTING METHODS AND PROPOSED MODEL

Method	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**False Alarm Rate (%)**
RNN	89.34	88.12	87.65	87.88	9.45
CNN	91.76	90.85	91.12	90.98	7.62
GAN	93.48	92.94	92.35	92.64	6.14
Proposed CAFMLN	97.92	97.45	98.12	97.78	2.36

Table 2 reveals that although RNN, CNN, and GAN models are characterized by a reasonable level of accuracy, they are characterized by an increased false alarm rate, as well as a lower level of balance between precision and recall. The proposed CAFMLN has the best accuracy of 97.92, as compared to all the approaches, with a false alarm rate of 2.36, and is more reliable and clinically effective in predicting early meningitis.

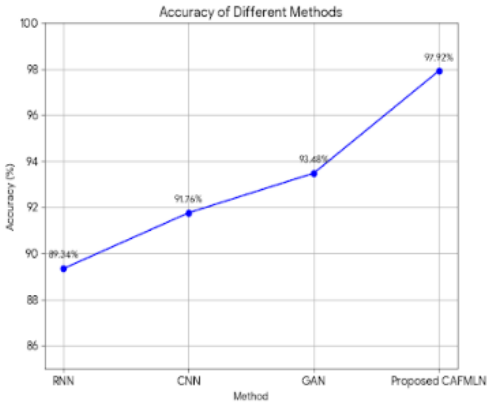

Fig. 3. Accuracy Comparison of Existing Methods and Proposed CAFMLN

Figure 3 shows the comparisons between model accuracies, as RNN, CNN, and GAN have 89.34%, 91.76 and 93.48, respectively, indicating the incremental feature learning. Compared to them, the proposed CAFMLN has much better performance of 97.92 and proves the power of causal-aware federated meta-learning and optimal choice of features to make sound predictions of meningitis in clinical decision support systems.

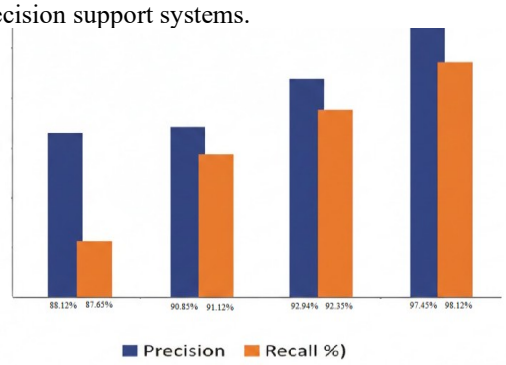

Fig. 4. Precision and Recall Comparison of Existing Methods and Proposed CAFMLN

Figure 4 estimates model precision and recall. RNN performs worse, whereas CNN and GAN exhibit gradual growth in their ability to identify a case of meningitis correctly. The proposed CAFMLN is the most accurate (97.45) and recalls the best (98.12), which proves its better capabilities of minimising misclassification and providing balanced and reliable clinical predictions.

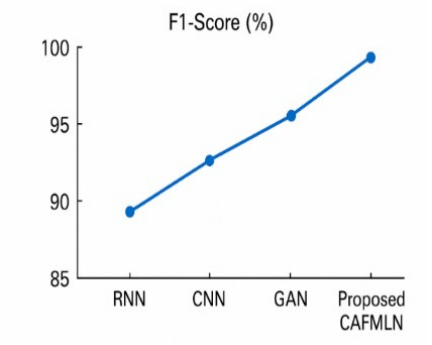

Fig. 5. F1-Score Comparison of Existing Methods and Proposed CAFMLN

Figure 5 shows the comparison of the F1-score of RNN, CNN, GAN, and CAFMLN models. The lowest balance is recorded by RNN and then better performance is achieved by CNN and GAN. The best F1-score of 97.78 of the proposed CAFMLN indicates its ability to provide a better balance of precision and recall, including a high reliability to predict clinical meningitis.

Fig. 6. Comparison of False Alarm Rate Across Methods

Figure 6 shows a comparison between RNN, CNN, GAN, and CAFMLN False Alarm Rates. The biggest decrease is registered by RNN (9.45%), then by CNN and GAN. The proposed CAFMLN has the lowest FAR of 2.36 percent, which shows that it is the least able to create false alerts and increase reliability in clinical prediction.

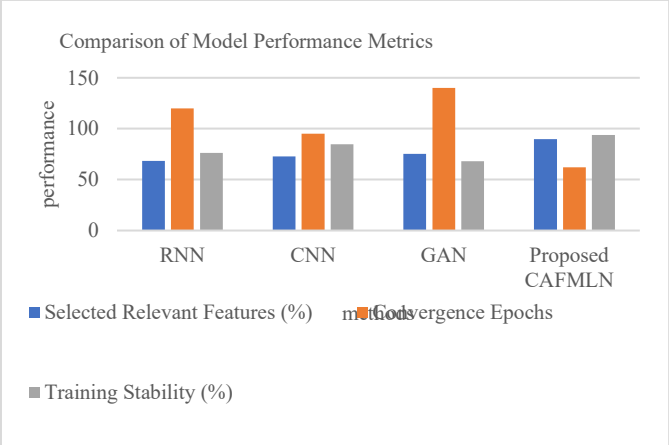

Fig. 7. Comparison of Model Performance Metrics: Feature Selection, Convergence, and Training Stability

The comparison of RNN, CNN, GAN, and CAFMLN, presented in Figure 7, is based on the feature selection, convergence epochs and training stability. CAFMLN is the best in terms of feature selection (89.72%), convergence (quickest, 62 epochs), and stability (93.86%). GAN has minimal stability and the lowest convergence speed, whilst CNN and RNN demonstrate intermediate efficiency and strengths, which proves that CAFMLN is more efficient and reliable.

TABLE IV. COMPONENT-WISE ABLATION ANALYSIS OF THE PROPOSED FRAMEWORK

Model Variant	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Without ADH	92.14	91.62	91.08	91.35
Without DCIS	91.48	90.95	90.27	90.61
Without ABHAO	93.02	92.44	91.96	92.19
Without NSCTN	90.76	90.12	89.58	89.84
Full Model (CAFMLN)	**95.63**	**95.08**	**94.72**	**94.90**

From the analysis in Table 4 above, it is clear that excluding each component reduces the model's performance. The model performs poorly without the NSCTN component, underscoring its importance for creating deep features. Likewise, without the DCIS component, the performance is poor since contextual feature learning is affected. The ADH component enhances adaptive feature learning, while the ABHAO component helps optimise attention. Both have a positive impact on the model's performance. The analysis shows that the complete model provides the best results for accuracy, precision, recall, and F1-score.

V. CONCLUSION

In conclusion, the suggested AI-based model is an excellent solution to early and successful meningitis detection, eliminating the issue of conventional diagnostics. The model brings together NSCTN and ADH, DCIS and ABHAO to capture the complicated clinical interactions, in addition to providing rapid convergence and high stability. The CAFMLN has 97.92 %, 97.45 %, 98.12 %, 97.78 %, and 97.92 % accuracy, preciseness, recall, F1-score, and false alarms respectively, which are impressive robustness and generalisation. The framework can be used with heterogeneous clinical data due to the effective feature selection and the stability of training. Subsequent investigations will involve integrating multimodal and lifelong learning, and lightweight deployable and understandable in under-resource healthcare settings.

REFERENCES

[1] Khodayar, Mahdi, Jianhui Wang, and Mohammad Manthouri. "Interval deep generative neural network for wind speed forecasting." IEEE Transactions on Smart Grid 10, no. 4 (2018): 3974-3989.

[2] Wang, Dong, Yuan Yuan, and Qi Wang. "Early action prediction with generative adversarial networks." IEEE Access 7 (2019): 35795-35804.

[3] Yang, Min, Junhao Liu, Lei Chen, Zhou Zhao, Xiaojun Chen, and Ying Shen. "An advanced deep generative framework for temporal link prediction in dynamic networks." IEEE transactions on cybernetics 50, no. 12 (2019): 4946-4957.

[4] Apalak, Merve, and Kamran Kiasaleh. "Improving sepsis prediction performance using conditional recurrent adversarial networks." IEEE Access 10 (2022): 134466-134476.

[5] Kadri, Farid, Abdelkader Dairi, Fouzi Harrou, and Ying Sun. "Towards accurate prediction of patient length of stay at emergency department: a GAN-driven deep learning framework." Journal of Ambient Intelligence and Humanized Computing 14, no. 9 (2023): 11481-11495.

[6] Zheng, Xiangtian, Bin Wang, Dileep Kalathil, and Le Xie. "Generative adversarial networks-based synthetic PMU data creation for improved event classification." IEEE Open Access Journal of Power and Energy 8 (2021): 68-76.

[7] Berghout, Tarek, Mohamed Benbouzid, Yassine Amirat, and Gang Yao. "Lithium-ion battery state of health prediction with a robust collaborative augmented hidden layer feedforward neural network approach." IEEE Transactions on Transportation Electrification 9, no. 3 (2023): 4492-4502.

[8] Song, Jingkuan, Yuyu Guo, Lianli Gao, Xuelong Li, Alan Hanjalic, and Heng Tao Shen. "From deterministic to generative: Multimodal stochastic RNNs for video captioning." IEEE transactions on neural networks and learning systems 30, no. 10 (2018): 3047-3058.

[9] Amanullah, M., and T. Kujani. "Advanced Meningitis Diagnostic Modeling: Combining Fuzzy Cognitive Maps and Decision Trees for Superior Predictive Performance." In 2025 International Conference on Emerging Technologies in Engineering Applications (ICETEA), pp. 1-6. IEEE, 2025.

[10] Doan, Bao Gia, Minhui Xue, Shiqing Ma, Ehsan Abbasnejad, and Damith C. Ranasinghe. "Tnt attacks! universal naturalistic adversarial patches against deep neural network systems." IEEE Transactions on Information Forensics and Security 17 (2022): 3816-3830.

[11] Zhao, Cong, Yifan Zhu, Yuchuan Du, Feixiong Liao, and Ching-Yao Chan. "A novel direct trajectory planning approach based on generative adversarial networks and rapidly-exploring random tree." IEEE Transactions on Intelligent Transportation Systems 23, no. 10 (2022): 17910-17921.

[12] Franchi, Gianni, Andrei Bursuc, Emanuel Aldea, Séverine Dubuisson, and Isabelle Bloch. "Encoding the latent posterior of bayesian neural networks for uncertainty quantification." IEEE Transactions on Pattern Analysis and Machine Intelligence 46, no. 4 (2023): 2027-2040.

[13] Bessadok, Alaa, Mohamed Ali Mahjoub, and Islem Rekik. "Graph neural networks in network neuroscience." IEEE Transactions on Pattern Analysis and Machine Intelligence 45, no. 5 (2022): 5833-5848.

[14] Habib, Ahsan, Mohammod Abdul Motin, Thomas Penzel, Marimuthu Palaniswami, John Yearwood, and Chandan Karmakar. "Performance of a convolutional neural network derived from PPG signal in classifying sleep stages." IEEE Transactions on Biomedical Engineering 70, no. 6 (2022): 1717-1728.

[15] Striuk, Oleksandr, and Yuriy P. Kondratenko. "Generative Adversarial Neural Networks and Deep Learning: Successful Cases and Advanced Approaches." Int. J. Comput. 20, no. 3 (2021): 339-349.

[16] Vital Jr, Adilson, Filipi N. Silva, Osvaldo N. Oliveira Jr, and Diego R. Amancio. "Predicting citation impact of research papers using GPT and other text embeddings." Physica A: Statistical Mechanics and its Applications (2025): 130789.

[17] Ivanovic, Boris, Karen Leung, Edward Schmerling, and Marco Pavone. "Multimodal deep generative models for trajectory prediction: A conditional variational autoencoder approach." IEEE Robotics and Automation Letters 6, no. 2 (2020): 295-302.

[18] Chava, Karthik, and Kanthety Sundeep Saradhi. "Emerging Applications of Generative AI and Deep Neural Networks in Modern Pharmaceutical Supply Chains: A Focus on Automated Insights and Decision-Making." (2024).

[19] Tao, Shengyu, Ruohan Guo, Jaewoong Lee, Scott Moura, Lluc Canals Casals, Shida Jiang, Junzhe Shi et al. "Immediate remaining capacity estimation of heterogeneous second-life lithium-ion batteries via deep generative transfer learning." Energy & Environmental Science 18, no. 15 (2025): 7413-7426.

[20] Errezgouny, Abderrachid, Youness Chater, Carlos D. Barranco González, and Abdeljabbar Cherkaoui. "An integrated deep learning approach for predictive vehicle maintenance." Decision Analytics Journal (2025): 100597.

[21] Hussain, Mohammad Kabir, Md Mustafizur Rahman, MD Shadman Soumik, Zunayeed Noor Alam, and Md Arifur Rahaman. "Applying Deep Learning and Generative AI in US Industrial Manufacturing: Fast-Tracking Prototyping, Managing Export Controls, and Enhancing IP Strategy." Journal of Business and Management Studies 7, no. 6 (2025): 24-38.

[22] Topşir, Aysel, Ferdi Güler, Ecesu Çetin, Mehmet Furkan Burak, and Melih Ağraz. "Thyroid disease classification using generative adversarial networks and Kolmogorov-Arnold network for three-class classification." BMC Medical Informatics and Decision Making 25, no. 1 (2025): 284.

[23] Becker, Marco, Philipp Niemietz, and Thomas Bergs. "Explainable neural network for time series-based condition monitoring in sheet metal shearing." Journal of Intelligent Manufacturing (2025): 1-17.