Rivera-Juzga, CY, Mistry, D, Knowles, AT, Lip, GYH, Ortega-Martorell, S and Olier, I

Graph neural networks for the enhanced prediction of new atrial fibrillation episodes after stroke

Computers in Biology and Medicine 197 (2025) 111095

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Graph neural networks for the enhanced prediction of new atrial fibrillation episodes after stroke

Cristian Y. Rivera-Juzga [c], Dharmesh Mistry [a,b], Adam T. Knowles [a,b], Gregory Y.H. Lip [b,d], Sandra Ortega-Martorell [a,b], Ivan Olier [a,b,*]

[a] Data Science Research Centre, Liverpool John Moores University, Liverpool, L3 3AF, UK
[b] Liverpool Centre for Cardiovascular Science at University of Liverpool, Liverpool John Moores University and Liverpool Heart & Chest Hospital, Liverpool, UK
[c] Escuela Colombiana de Ingeniería Julio Garavito, Bogotá, Colombia
[d] Danish Centre for Health Services Research, Department of Clinical Medicine, Aalborg University, Denmark

ARTICLE INFO

Keywords:
Graph neural networks
Risk prediction
Machine learning
Atrial fibrillation
Stroke
Intensive care unit
Clinical decision support

ABSTRACT

Traditionally, statistical and machine learning (ML) algorithms have been used to develop risk prediction models for adverse clinical events, such as atrial fibrillation (AF) after stroke. However, these algorithms often fail to encapsulate or exploit possible connections between patients, assuming each patient is fully independent. This study builds a graph of patients interlinked by their medical histories to create a graph-based risk prediction model for AF. We investigate the ability of Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT) to predict AF risk in critically ill stroke patients. We introduce a novel, patient-specific approach for computing similarities between GNN nodes and explore several methods for GNN explainability, including node-specific Shapley value analysis and node relationships based on the attention coefficients of the GAT model. Our findings show that GCN and GAT models, with AUCs of 0.81[0.78–0.84] and 0.84[0.81–0.87], consistently outperform traditional algorithms such as Random Forest, XGBoost, and Logistic Regression, which had the best AUC of 0.78 [0.74–0.82]. This superior performance is observed when our proposed custom similarity metric is used to construct the graph, highlighting the importance of task-specific graph design in enhancing model effectiveness. The attention mechanisms in GAT models likely contributed to this improved performance. This study highlights the strength of GNNs in capturing complex relationships and provides insights into model predictions, demonstrating the generalisability of our methodological approach to other risk prediction models.

1. Introduction

Stroke is a serious medical condition that occurs when blood flow to the brain is interrupted, resulting in brain damage, and potentially leading to permanent disability or death [1]. Currently, more than 101 million people live with the effects of stroke; a figure that has doubled during the last 30 years, with around 63 % of cases occurring to people under 70 years of age [2]. Many patients who present with a stroke (both ischaemic or haemorrhagic) are at high risk of mortality, and morbidity, including from incident cardiovascular events, the so-called stroke heart syndrome [3–5].

Atrial fibrillation (AF) is the commonest sustained cardia rhythm disorder and is an important predisposing factor for ischaemic stroke [6]. AF affects over 2 % of the European population and accounts for approximately 59 million cases globally [7]. Stroke is usually more severe when it is related to AF, is often more severe, making the identification of AF risk after a stroke crucial for preventing recurrence [8]. Notably, an estimated 1.5 million new cases of AF occur annually following a stroke [9,10]. While around 20 % of ischemic stroke patients have a pre-existing AF diagnosis, an additional 10–20 % of patients may be diagnosed with AF after undergoing cardiac monitoring [10].

In 10 % of patients with ischaemic stroke, new AF is detected after standard diagnostic workup, and up to 10 % could be diagnosed with extensive workup, including in-hospital monitoring or ambulatory monitoring for several days in most hospitals [11,12]. However, a substantial proportion of patients with AF will still remain undiagnosed, posing them with a higher risk of recurrent stroke, and opportunities for targeted management, including thromboprophylaxis, will potentially be lost. Hence the importance of identifying patients at risk of developing AF after suffering from a stroke.

* Corresponding author. Data Science Research Centre, Liverpool John Moores University, Liverpool, L3 3AF, UK.
E-mail address: I.Olier@ljmu.ac.uk (I. Olier).

https://doi.org/10.1016/j.compbiomed.2025.111095
Received 28 December 2024; Received in revised form 24 June 2025; Accepted 12 September 2025
Available online 17 September 2025

Traditionally, statistical methods (e.g. Logistic Regression) and, more recently, machine learning (ML) algorithms have been used for the development of risk prediction models of adverse clinical events, such as AF after stroke (e.g. Refs. [13,14]). However, such algorithms often do not properly encapsulate or exploit possible connections between patients (i.e. patients sharing similar characteristics), often treating stroke risk factors in isolation or in a binary (yes/no) manner. Also, these algorithms often assume that every patient is fully independent of others. In this work, we take the approach of building a graph of patients, under the assumption that stroke patients have some commonality with their past medical histories, to develop a graph-based risk prediction model of AF post-stroke.

Graphs are a common way to analyse relationships between objects. As such, they are used for numerous real-world problems, e.g. social media [15], bioinformatics [16], epidemiology [17], network analysis [18] and computer vision [19]. The unique ability of graphs is that structural relationships between data can be obtained, which can allow for more information to be gained than through analysis of the data in isolation. Despite this, it is often very difficult to solve learning problems in graphs, partly because many types of data are not typically structured as graphs and secondly, the underlying connectivity patterns of graph-structured data are often complex and diverse [20].

Graph Neural Networks (GNNs) belong to the family of ML algorithms capable of processing graph data, allowing for the discovery and modelling of complex interactions between multiple factors and learning from the relationships between nodes. Unlike conventional models, GNNs can identify and weight associations between various factors and how these interrelationships can influence a given outcome [21]. Although several GNN architectures have been proposed, perhaps the most popular ones are Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs). GCNs have been proposed to solve several different problems [22]. Compared to other models, GATs can learn the information from the neighbours of a node, by weighting each neighbour according to its relevance to the node of interest [23]. This approach allows GATs to be a more flexible model with a higher level of interpretability, as attention coefficients can be considered to indicate the importance of each neighbour, and they have been used in various applications, with several reviews available [24,25].

GNNs have been implemented in the risk prediction of outcomes of patients in and emergency and critical care, particularly to exploit the use of time series, diagnosis and medical data simultaneously, in addition to the relational data found by connecting patients with similarities [26,27]. For example, Tong et al. [26] approached the prediction of patient in-hospital mortality and length of intensive care unit (ICU) stay through the combination of a Long-Short Term Memory (LSTM) network and a GNN that could use both ICU time series and neighbouring patient data. It was shown that the additional use of data from patients with similar characteristics allowed for a better performance of outcome prediction compared to models that learned from time series data alone. In the study by Xu et al. [28], the authors also developed a predictive decision-support model of two typical ICU interventions, again using temporal and spatial (through graph representation) patient data in tandem, focusing on models that can be clinically interpreted and therefore used by medical practitioners. The hybrid approach was also demonstrated as an improvement in performance, compared to a variety of traditional ML methods, LSTM and CNNs.

Leveraging recent developments in GNNs, we aimed to investigate their ability to predict AF in critically ill stroke patients based on data from the Medical Information Mart for Intensive Care (MIMIC)-IV database [29]. During this process, we define a novel approach for computing the similarities between GNN nodes that is more instance (i.e. patient) specific than the popular cosine similarity method. We also explore several methods for GNN explainability, specifically node specific Shapley value analysis, and node relationships based on attention coefficients of the GAT model, thus highlighting the strength of GNNs in capturing complex relationships and providing insights into how our models have produced their predictions. The performance of the resulting GCN and GAT models are compared with the more traditional algorithms; Random Forest, XGBoost, and Logistic Regression. Furthermore, the methodological approach employed in this work is not limited to AF prediction alone but is generalisable to many applications of GNNs.

2. Methods

2.1. Dataset extraction

Data was extracted from the MIMIC-IV database (version 2.2), which houses from than 200,000 patients from Boston, MA, USA [29]. Only those patients diagnosed with stroke in their admission evaluation or during their hospital stay were included in this study. For each of these patients, we obtained the following data: records of medical diagnoses and evaluations (using the WHO's International Classification of Diseases version 10, ICD-10); demographic information; averages of vitals and blood tests taken during the first 24 h of their ICU stays; in addition to the presence of AF after their first 24 h (model outcome). This was supplemented by patient history information, namely the number of prior hospital admission for stroke or otherwise, and the number of days since their last admission. This data forms a table where each row is a patient, and each column pertains to one of the above data types (or variables). Patients with AF episodes during the first 24 h after ICU admission were excluded. Data pre-processing details are included in the Supplementary Materials.

2.2. Machine learning algorithms

For the classification task of predicting the risk of developing AF, we evaluate the of use of traditional ML algorithms as well as GNNs. Details of the hyperparameter selection are included in the Supplementary Materials.

2.2.1. Traditional algorithms

The traditional algorithms evaluated were Logistic Regression, XGBoost and Random Forest. We use them as a comparative basis for the assessment of GNNs. Logistic regression, used for binary classification, employs the logistic (sigmoid) function to transform a linear combination of inputs into the probability of a binary outcome, which can then be thresholded for class prediction. XGBoost is an efficient implementation of gradient boosted decision trees, which builds an ensemble of trees sequentially to correct the errors of the previous ones. It uses regularisation and parallelisation to prevent overfitting and improve accuracy. Random Forest is another ensemble learning method that combines the predictions of multiple decision trees through majority voting and each tree is trained on random subsets of data and features (for classification) to reduce variance and control overfitting.

2.2.2. Graph neural networks (GNNs)

GNNs are designed to process data structured as graphs consisting of nodes (representing data instances with defining features) and edges (representing the relationships between instances). The basic principle of GNNs is to learn from the relationships between nodes through a process of passing information along the edges connecting them. In this work, we evaluate the use of GCN and GAT models for prediction modelling.

The operation of a GCN, whose fundamental functions are graph convolutions, is based on learning through an embedding and weighted aggregation process of each node in a graph, using the information of adjacent nodes and characteristics of the nodes themselves. The convolution weights and combines node and edge characteristics from neighbouring nodes to generate new representations of each node. In general terms, a GCN is similar to a Convolutional Neural Network (CNN), where each convolution layer extracts the most relevant

characteristics of the node and transmits this information to the next layer until it has a complete learning of the data. The process is repeated across multiple convolutional layers to capture increasingly complex patterns. In this work, we follow the general representation and methodology of a multi-layer GCN described in Ref. [30].

GATs extend the functionality of GCNs by introducing an attention mechanism. This allows the model to assign different levels of importance (captured by attention coefficients) to each neighbour of a given node. The computed attention coefficients reflect the relevance of neighbours to the target node, are normalised in the vicinity of each node and are used to weight the features of neighbours before their aggregation. The importance scores in each iteration are then normalised in the vicinity of each node using a Softmax function and the importance of the neighbourhood in the model is calculated by a process of self-attention between each pair of nodes. This approach facilitates more nuanced and flexible node representations, as GATs can focus on the most relevant neighbours. The attention mechanism aids the interpretability of the model, as the attention coefficients provide details of the importance of each neighbour [23]. In this study, we followed the generalised architecture and approach outlined in Ref. [23].

2.3. Graph construction

The GNNs require a graphical representation of data training inputs. While the node information is typically presented as a table of patients and their associated features, the edges are represented by way of an adjacency matrix, where some similarity measure of the connection between every pair of nodes is stored. The following details how we define the nodes and edges and compute the similarities for the adjacency matrix.

2.3.1. Nodes and edges

Each node in the graph represents a patient along with their associated patient data, which consists of variable values, apart from the AF target variable. The connection between nodes, i.e. edges, are represented by the similarity of diagnoses between patients. This serves as a measure of the relationship between nodes.

Each patient's diagnoses form a list of ICD codes, where each code represents a particular condition, for example, codes I63.432, R47.01, and R13.10, represent cerebral infarction, aphasia, and dysphagia, respectively. To calculate the similarity between patients, it is first necessary to one-hot-encode the diagnoses.

The calculation of similarity coefficients between patients that form the adjacency matrix is performed using two different methodologies, which are assessed in terms of model performance as part of the model optimisation stage.

2.3.2. Cosine similarity

For the first method, we compute the cosine similarity between two vectors in diagnosis space using the following equation:

$$cos(\theta) = \frac{\boldsymbol{A}\cdot\boldsymbol{B}}{\|\boldsymbol{A}\|\cdot\|\boldsymbol{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} A_i^2}} \quad (1)$$

Here $\boldsymbol{A}$ and $\boldsymbol{B}$ are one-hot encoded vectors representing the diagnoses of two different patients, where each element in the vector indicates the presence or absence of one of n specific ICD codes (n=*4129*).

2.3.3. Custom similarity measure

The second methodology is a novel method we propose in this work and can be represented through the following set of equations.

$$Z_j \begin{cases} 1 & \text{for } A_j = 1, B_j = 1 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$$\text{Similarity} = \frac{\sum_{j=0}^{n-1} Z_j}{\max(\|\boldsymbol{A}\|, \|\boldsymbol{B}\|)} \quad (3)$$

Z is the vector formed by comparing each position j in the diagnosis vectors **A** and **B** and setting each element Z_j to 1 if the diagnoses match and 0 otherwise. The similarity coefficient is then calculated by dividing the sum of **Z** by the greater of the total number of diagnoses for the two patients, whose diagnoses are represented by **A** and **B**, respectively. This method considers diagnoses in common and the total number of diagnoses, where the latter accounts for the disparity in the overall number of diagnoses between patients.

In both cases, the result is a m x m adjacency matrix, where m is the number of patients analysed and each value at position (i, j) of this matrix corresponds to the similarity measure between patients according to their diagnoses.

In summary, our graph consists of nodes that represent patients, with node feature vectors incorporating patient data that included demographic information and 24-h averaged blood and vital tests. Edges between nodes were constructed based on a similarity measure between the one-hot encoded vectors of patients' conditions, in the form of ICD records of medical diagnoses, on admission. If the similarity between patients was above an optimised threshold, an undirected, unweighted edge was created between the nodes.

In our modelling approach, we optimise both the method of similarity measure and the total number of diagnoses considered during the hyperparameter tuning for the GAT and GCN models separately. Our graph construction and GNN model framework is shown in Fig. 1. To achieve this, we employed the Optuna hyperparameter optimisation package [31], which uses adaptive tree-structured Parzen estimators (TPE) to efficiently explore the hyperparameter space and identify the combination that yields the best model performance. This adaptive sampling strategy allows for a more intelligent and directed search, reducing computational cost while increasing the likelihood of finding an optimal configuration. Further implementation details are provided in Section 2.2 of the supplementary materials.

2.4. Evaluation and assessment methods

Evaluation of the models was performed using accuracy, recall (or sensitivity), precision, and area under the ROC Curve (AUC). For model interpretability purposes, we use SHAP (SHapley Additive exPlanations) plots (specifically, the beeswarm plot) for assessing variable (or feature) importance for the base models [32]. Feature importances for the GNN models are calculated using the GNNExplainer module of PyTorch [33]. For our GAT model, we construct a visualisation focusing on a particular node and signifying the attention weight by the thickness of the connection between nodes that are connected to it, to help understand the neighbouring nodes that are most influential in the model's attention mechanism [26]. As a final examination of the GAT model, the final attention layer, upon which predictions are based, is modelled with t-SNE [34]. This helps visualise the effectiveness of the attention framework in separating patients with and without AF.

3. Results

3.1. Resulting dataset

The extracted dataset comprised 1949 stroke patients, with an imbalanced frequency distribution such that 12 % of patients had an episode of AF after 24 h of ICU admission, while the remaining 88 % did not. Putting aside the medical diagnoses and the outcome variable, more than 37 % of the variables had a considerable amount of missing data, where missingness was as high as 98 %.

Table 1 lists the variables that were used for modelling following the pre-processing stages. The mean, standard deviation and interquartile

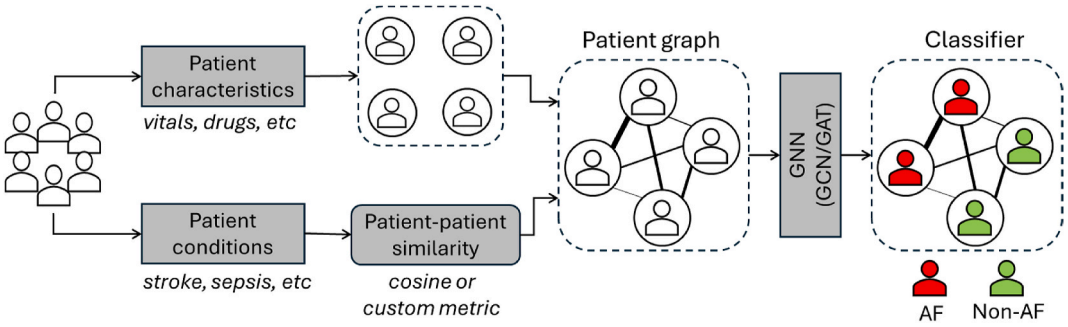

Fig. 1. Overall methodological framework. The modelling data are structured as a patient graph. Patient characteristics such as vital signs, demographics and drugs constitute the graph nodes, while the edges represent patient-to-patient similarities based on the number of shared medical conditions. GNNs are employed to predict whether patients are at risk of developing AF.

Table 1
Variables selected for model training (mean ± standard deviation and inter-quartile ranges are provided for continuous variables; and frequencies and proportions for categorical ones). GCS: Glasgow Coma Score; Stroke Diagnosis Priority: importance given in the diagnosis, with 1 and 4 representing the highest and lowest levels of priority, respectively; Blood Pressure: mean of Systolic and Diastolic Blood Pressure.

Variable	Characteristics
AF (Target Variable) [Yes]	239 (12.3 %)
Age (years)	66.5 ± 15.4 [57–78]
Sex Male [Yes]	1023 (52 %)
Stroke Diagnosis Priority (number)	3.4 ± 4.6 [1.0–4.0]
Previous Stroke Admissions (number)	0.1 ± 0.4 [0.0–0.0]
Previous Other Admissions (number)	0.6 ± 1.9 [0.0–0.0]
Days Since Previous Admission (number)	500.4 ± 710.7 [31.5–713.0]
GCS Eye Response[a]	3.0 ± 1.0 [2.4–4.0]
GCS Motor Response[a]	5.1 ± 1.4 [4.5–6.0]
GCS Verbal Response[a]	3.2 ± 1.7 [1.0–5.0]
Temperature[a] (°C)	36.9 ± 0.5 [36.7–37.1]
Diastolic Blood Pressure[a] (mmHg)	65.5 ± 12.8 [55.6–73.9]
Systolic Blood Pressure[a] (mmHg)	128.4 ± 18.7 [114.3–141.4]
Blood Pressure[a] (mmHg)	83.3 ± 12.74 [73.5–91.9]
Heart Rate[a] (beats per min)	79.2 ± 14.1 [69.4–88.1]
Respiratory Rate[a] (breaths per min)	18.9 ± 3.6 [16.4–20.7]
Oxygen Saturation[a] (%)	96.6 ± 2.5 [95.6–98.3]
Haemoglobin[a] (g/dL)	11.3 ± 2.1 [9.8–12.9]
Glucose[a] (mg/dL)	139.0 ± 44.8 [109.7–155.8]
Anion Gap[a] (mEq/L)	14.0 ± 3.3 [12.0–16.0]
Creatine[a] (mg/dL)	1.2 ± 1.3 [0.7–1.2]
Magnesium[a] (mg/dL)	2.1 ± 0.4 [1.8–2.2]
Phosphate[a] (mg/dL)	3.5 ± 1.0 [2.9–3.9]
Platelet Count[a] (K/uL)	213.6 ± 95.4 [152.6–257.0]
Potassium[a] (mEq/L)	4.1 ± 0.5 [3.8–4.4]
Prothrombin Time[a] (seconds)	13.8 ± 4.0 [11.9–14.4]
Weight[a] (kg)	80.0 ± 21.0 [65.9–91.9]

[a] For these variables, the values used are the mean averages of the associated measurements.

ranges are provided for continuous variables, and frequencies and proportions for the categorical ones. The medical diagnoses of each patient in the form of ICD codes are supplied as complementary information. The codes form a list, the length of which is determined by the number of medical diagnoses the patient has. These diagnoses form part of the GNN input. There are 4129 different diagnoses, and each patient may be diagnosed with multiple conditions, therefore a matrix of 1949 rows (one per patient) and 4129 columns (one per diagnosis) is the result. The data were randomly split into training and test sets in a 75:25 ratio, with both sets maintaining the same prevalence of AF as the full dataset.

3.2. Base model optimal hyperparameters

For the base models, hyperparameter optimisation was performed using a 5-fold cross validation, with Table 2 showing the optimal hyperparameters found.

Table 2
Optimal hyperparameters for traditional models.

Algorithm	Hyperparameter	Value
Logistic Regression	Penalty	L2
	C	100
XGBoost	Maximum Depth	2
	Number of Trees	200
	Learning Rate	0.01
Random Forest	Number of trees	100
	Max Features	sqrt
	Max depth	4
	Criterion	gini

3.3. GNN model optimal hyperparameters

With the aid of the Optuna package, the optimisation of the GNN models was performed in two stages, the optimisation of the graph layout and the optimisation of the GNN architecture.

3.3.1. Graph optimisation

The optimal graph properties for the GCN and GAT models are shown in Table 3. For both GCN and GAT, that the optimal methodology for computing the similarities between nodes was found to be our custom-made approach. The number of diagnoses considered was significantly reduced, while at the same time around 80 % of the node linkages were retained, which considerably reduces the computational complexity.

3.3.2. GNN model optimisation

Optimal hyperparameters are given in Table 4. We see that optimal model architectures are not required to be particularly complex, with networks less than 4 layers deep and a drop-out fractions close to 0.5.

3.4. Model results

Table 5 shows the performance of each of our evaluated AF risk prediction models as measured using the test subset. ROC curves for each model are given in Fig. 2. In Tables S2 and S3 of the Supplementary Materials, we also provide the confusion matrices for the optimised

Table 3
Optimal Graph properties for GNN models.

Model	Diagnosis Matrix Method	Number of Diagnoses	Similarity Threshold
GCN	custom	446	0.205
GAT	custom	313	0.218

Table 4
Optimal GNN model hyperparameters.

Hyperparameters	GCN Model Values	GAT Model Values
Number of Layers	4	2
Number of Hidden Channels	57	56
Drop Out	0.43790	0.60081
Learning Rate	0.02175	0.01308
Decay	0.00004	0.00003
In Head	–	3
Out Head	–	10

traditional ML and GNN risk prediction models, respectively.

The Random Forest model produces the strongest overall performance amongst the traditional (or base) models, producing the highest accuracy, recall and precision scores, and AUC similar to the highest scoring base model (Logistic Regression). However, the results indicate that the GNN models outperformed the base ML models across most performance metrics. The GAT model achieved the highest accuracy at 0.80, significantly surpassing the best base model for this metric. Moreover, the GNN models demonstrated enhanced recall and precision, with GAT leading in recall at 0.73 and in precision at 0.35. In terms of AUC, both GNN models excelled, with GCN reaching 0.81 and GAT at 0.83, both higher than the best base model, showing the effectiveness of GNN models to capture complex relationships and dependencies between nodes. When comparing F1-scores, the GAT performs best, with a score of 0.47 compared to the next best models of GCN and Random Forest that both achieve a score of 0.43. In specificity, both GAT and GCN models also lead in performance, achieving scores of 0.81 and 0.78, respectively, compared to 0.75 values in all the base ML models.

Both GNN models achieved better performance with our custom-defined similarity measure, outperforming the cosine similarity method in every metric provided. Specifically, the GCN model saw an improvement in AUC from 0.76 to 0.81, while the GAT model improved from 0.77 to 0.83. The improvements from using the custom over cosine method are generally more pronounced in the GAT model, which can be attributed to its dynamic attention mechanism. The GAT model is more likely to benefit from improvements in the connections between nodes, compared to a GCN model that uses a simpler method of neighbour information aggregation.

In summary, the GNN models, particularly those using our custom-defined similarity method, demonstrate advantages over the base ML models. Specifically, the GAT [custom] model performed best across all metrics given in Table 5, achieving the highest AUC, accuracy, recall, precision, F1-score and specificity. The GCN [custom] model also showed improvements over traditional methods, particularly in recall and F1-score. While the Random Forest model performed well among the base models, it did not surpass the performance of the GNN models, highlighting the potential of GNNs to capture more complex relationships and patterns within the data by connecting patients (nodes) with similar conditions.

3.5. Feature importances

3.5.1. Base models

The feature importances, in the form of beeswarm SHAP plots, are obtained for the best performing base model, Random Forest, shown in Fig. 3.

The Glasgow Coma Scale (GCS) [35] is a clinical tool used to assess a patient's level of consciousness, based on eye-opening, verbal response and motor response. They appear to be significant variables in the prediction of AF according to Fig. 3. The mean GCS motor response was found to have the greatest relevance; the result is supported by a study

Table 5
Performance metrics for ML models. AUC 95 % confidence intervals are given in brackets.

Model	AUC (95 % CI)	Accuracy	Recall	Precision	F1-score	Specificity
Logistic Regression	0.77 [0.73–0.81]	0.74	0.65	0.28	0.39	0.75
XGBoost	0.76 [0.72–0.80]	0.73	0.63	0.27	0.38	0.75
Random Forest	0.78 [0.74–0.82]	0.75	0.71	0.31	0.43	0.75
GCN [cosine]	0.76 [0.71–0.79]	0.75	0.67	0.28	0.40	0.76
GCN [custom]	0.81 [0.78–0.84]	0.77	0.71	0.31	0.43	0.78
GAT [cosine]	0.77[0.73–0.81]	0.72	0.72	0.26	0.38	0.72
GAT [custom]	0.83 [0.81–0.87]	0.80	0.73	0.35	0.47	0.81

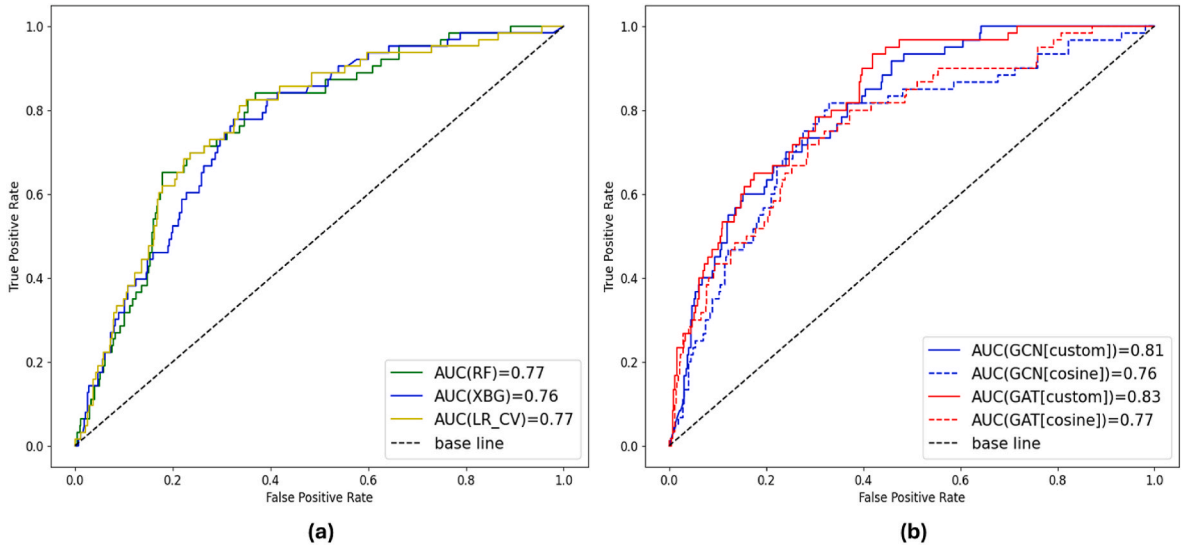

Fig. 2. AUC ROC Curves for each ML model. a) The labels RF, XGB, LR_CV refer to the Random Forest, XGBoost and Logistic Regression models, respectively. b) For GNNs with a similarity method in the graph construction given by the label in the square brackets.

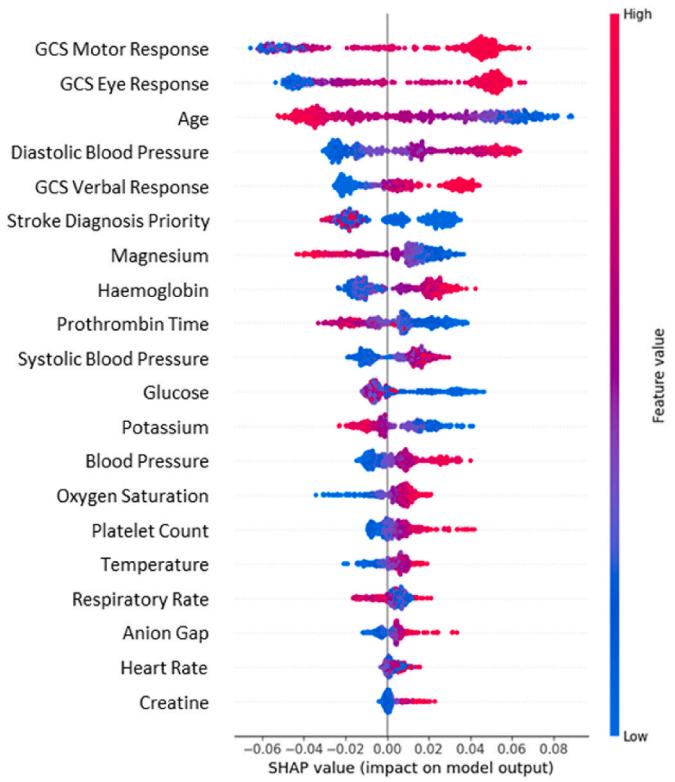

Fig. 3. Beeswarm SHAP plot based on SHAP values for the Random Forest model.

that found a significant association between a low GCS score and the presence of AF, suggesting a decline in the level of consciousness may be related to an increased risk of AF [36]. We can see this aspect in the distribution, higher values of the GCS score correspond to positive SHAP values, i.e., more likely to be predicted as not having AF.

Patient age is deemed significant in the prediction of AF, supported medically with links between increased age and the development of AF [37]. Again, the distribution of ages in Fig. 3 reflects this, higher ages are associated with negative SHAP values, i.e., higher likelihood of developing AF. Diastolic blood pressure is another variable deemed influential in AF prediction, appearing within the top 4 of the feature importances, again in agreement with medical findings of links between elevated diastolic blood pressure and increased risk of developing AF [38].

3.5.2. GCN model

For the GNN, feature importances are computed at node level based on calculations of mutual information for the subgraph associated with the selected node [39]. We obtain the feature importances of patients indexed at rows 218 and 1643, which were selected at random from the dataset. These are shown in Fig. 4.

In Fig. S1 of the Supplementary Materials we can make a comparison of feature importances between the two selected nodes (patients). What is evident is that the level of importance a particular feature has for AF prediction differs between the two patients. Differences in importances from patient to patient are reasonable to assume, given that no two patients are identical. We can further emphasise this with Fig. 4, in which the features are arranged by feature importances. Most relevant for the AF prediction of patient 218 is the stroke diagnosis priority, while for patient 1643 it is the number of previous admissions for stroke. Such differences in the order of relevance are present as we look further down each patient's ordered importance list (e.g., see relative positions of sex, heart rate and weight). These node specific importances represent an advantage over the overall measures used in traditional models, providing a level of interpretation that would be particularly advantageous in a clinical setting where patient level understanding is key.

Examining the feature importances in general, the number of days since the last hospital admission and the number of previous admissions due to ischaemic stroke are particularly relevant. The latter is supported by various medical studies that have shown and discussed the inverse relationship between ischemic stroke and AF in the patient, i.e., the presence of AF in the patient increases the risk of the patient presenting a scenario of ischemic stroke [40–42]. Studies have also shown an association between hyperglycaemia and the risk of developing AF [43,44]; a relation also reflected in GCN model that found that the mean glucose level was a significant variable for its AF prediction.

3.5.3. GAT model

The feature importances for the GAT model, based on the same method as for the GCN, are shown in Fig. S2 of the Supplementary Materials. What is immediately clear is that there is little difference in feature importance between each of the variables in comparison to the GCN where clear differences exist. One may attribute this to the GAT model assigning a higher priority to the relationship that exists between neighbouring nodes and their connection strength than the GCN model, that may focus more on the node features.

Comparing the importances between the same two patients (Fig. S2) one can see that importances are higher for all variables for patient

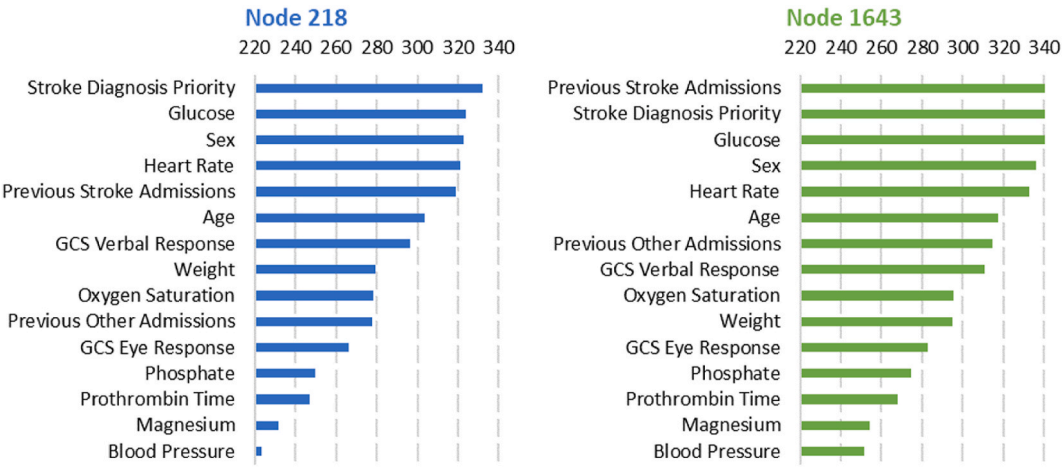

Fig. 4. Feature importances for nodes 218 and 1643 in the GCN model, showing the top 15 most important features per node, organised in descending order of importance.

1643. However, only through analysis of Fig. 5 may we identify any difference between the patients. Differences in the order of relevance are present, e.g., diastolic blood pressure and prothrombin time, though with differences in importance scores minor for each patient, the significance of the ordering can be deemed minor. This merely reflects what is mentioned in the previous paragraph.

Despite this, it is evident that the mean of the respiratory rate is significant in AF prediction for the GAT model. This aligns with studies [45,46] showing that the presence of AF in patients with elevated respiratory frequency is associated with an increased risk of developing cardiovascular complications and cerebrovascular events. Furthermore, as with the GCN, the GAT places significant importance to historical patient data, e.g., days since last hospital admission and the number of admissions due to ischaemic stroke.

3.6. GAT attention visualisations

Fig. 6 displays the five most relevant nodes to node 218 according to attention coefficients. Based on the thickness of the connections (corresponding to the attention coefficients), node 1389 holds the greatest relevance to node 218, followed by nodes 875, 568, 1329 and 1277. By examining the nodes further, we may understand the reasons behind their relevance. Therefore, in addition to Fig. 6 we provide Table 6, which provides several of the variables for patients associated with each node.

The top three rows of Table 6 list patients diagnosed with AF, showing that their blood pressure, haemoglobin, and diastolic blood pressure are lower than those of the patients in the last three rows (non-AF). Additionally, the respiratory rate for AF patients is higher compared to non-AF patients. This pattern indicates that these variables—blood pressure, haemoglobin, diastolic blood pressure, and respiratory rate—play a significant role in predicting AF for node 218. The GCS verbal response variable values tend to lie within a range of 4–5, while the interquartile range of the variable considering all patients is 1–5. The GAT model, therefore, may consider values with the range 4–5 significant in identifying nodes most relevant to 218. This relevancy is further emphasised for the haemoglobin variable, where the most relevant node, 1389, has a very similar value to node 218.

To conclude our GAT attention visualisation, Fig. 7 provides the reduced dimensionality representation of the node embeddings of the final attention layer of the GAT model. The dimensionality reduction is performed with the scikit-learn implementation of t-SNE using the recommended perplexity value of 30. It is evident that the model tends to group patients (nodes) by the existence of AF, thus learning from the patient characteristics and connections to ultimately lead to a strong predictive performance.

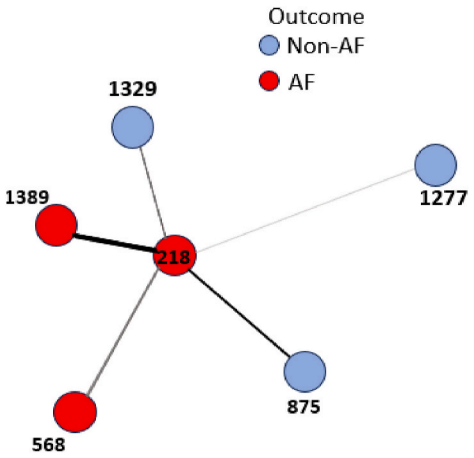

Fig. 6. Attention visualisation for node 218. Displayed are the five most relevant nodes linked to node 218 based on attentions scores. The thicker the line connecting to node 218, the higher the attention coefficient, and in turn, the higher the relevance of the connecting node. Blue coloured nodes denote patients without AF, while those coloured red denote patients diagnosed with AF.

4. Discussion

Understanding the risk of AF in stroke patients is critical for effective prevention and management of recurrent strokes, and the integration of advanced ML technologies offers transformative potential by enhancing risk stratification and early detection [47]. In this study, we present a novel application of graph neural networks (GNN) for the prediction of AF after stroke. Our findings highlight the significant advantages of GNNs, demonstrating their superiority over traditional machine learning and biostatistical methods in this context.

The use of GNN was based on patient medical diagnoses, ICU data, and demographic data, and the performance analysis reveals a clear advantage of GNNs over traditional ML models in the task of node classification. We compared the performance of traditional ML models (Logistic Regression, Random Forest, and XGBoost) to more advanced neural network models in the form of GNNs, more specifically GCNs and GATs. We proposed a novel custom similarity measure for defining the links between nodes in the graphs that serve as GNN inputs. Furthermore, we demonstrated methods for interpretability of the both the traditional and GNN based models.

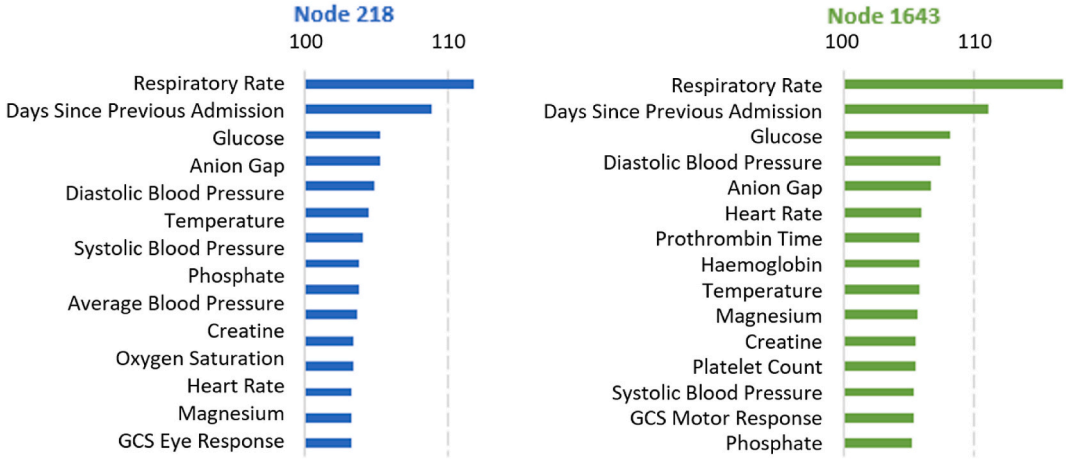

Fig. 5. Feature importances for nodes 218 and 1643 in the GAT model, showing the top 15 most important features per node, organised in descending order of importance.

Table 6
Characteristics of patients displayed in Fig. 5.

Node (Patient)	Outcome	Blood Pressure (mmHg)	Respiratory Rate (breaths/min)	Haemoglobin (g/dL)	Diastolic Blood Pressure (mmHg)	GCS Verbal Response
218	AF	76.50	21.65	7.67	56.00	4.17
1389	AF	65.75	17.67	7.78	45.23	5.00
568	AF	64.21	26.53	10.33	50.94	4.67
1329	Non-AF	88.25	17.60	15.40	76.27	4.00
1277	Non-AF	86.25	14.63	13.10	73.38	5.00
875	Non-AF	98.96	16.23	13.80	83.83	5.00

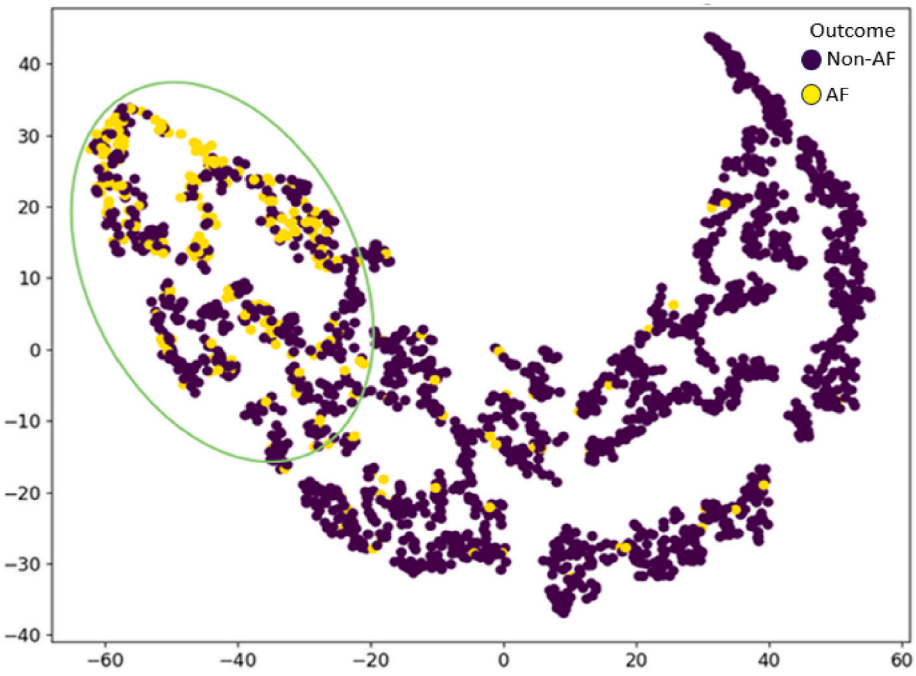

Fig. 7. t-SNE reduced dimensionality representation of the node embeddings of the final attention layer of the GAT model. Patients (nodes) detected with AF are coloured in yellow, while those without AF detection are coloured purple.

While the Random Forest model stands out among the base models, delivering the highest accuracy, recall, and precision scores, its performance is ultimately surpassed by the GNN models. Specifically, the GAT model achieves the highest performance overall (AUC: 0.83 [0.81–0.87]; accuracy: 0.80; recall: 0.73; precision: 0.35) a notable improvement over the best performing base model. This enhanced performance highlights the GAT's superior ability to capture intricate patterns and relationships within the data and between examples, which traditional models may find harder to identify. Although we did not directly assess the effect of the GAT's attention mechanisms on the models, we believe that the role they play could explain their observed higher performance.

The differences in results found in Figs. 3–5 can attributed to several factors. Firstly, Fig. 3 shows the aggregated SHAP values from the Random Forest model across all patients in dataset, whereas the GNN results in Figs. 4 and 5 present SHAP values for two specific patients (nodes 218 and 1643), providing a more individualised interpretation. Secondly, the differences observed in Figs. 4 and 5 highlight fundamental distinctions between how GCN and GAT models aggregate and weigh node features. Unlike GCN, GAT models employ attention-based aggregation, learning individualised attention weights for neighbouring nodes based on their relative importance. This dynamic, learned weighting naturally leads to a distance – and often more nuanced – distribution of SHAP values. As a results, when GCN and GAT models are trained with independently optimised hyperparameters or graph structures (e.g., varying the optimal number of diagnoses or similarity metrics), as in our study, the resulting SHAP values will inherently differ due to variations in structural definitions and feature relationships. It is important to emphasise here that the primary objective of our study was hyperparameter and graph optimisation for model performance comparison, rather than on interpretability.

Direct comparisons of our model performances with other works that predict AF may be limited by differences in databases used, patients, input data types, and the specific ML algorithms used amongst other factors. To highlight this point, there are numerous studies outlined in our review [48] that report the application of ML for AF modelling, although reported scores are mostly for AF detection rather than risk prediction. Also, electrocardiograms (ECGs) were used in most studies with input either directly or following some transformation. AF detection is a significantly easier task than risk prediction. AUC scores range from 0.92, achieved by Ref. [49] using ECGs transformed with wavelet transform as input for a convolutional neural network, to 1.00 by Faust et al. [50] though with only 23 ECG records directly input into a LSTM network. More relevant work for comparison would be our prior study [51], which presents a logistic regression model for the prediction of AF in critical care patients, though using data (AmsterdamUMCdb) and a different study design (not specific for stroke patients). Model AUC scores were 0.82 for ventilated patients and 0.91 for non-ventilated cohorts. However, one must emphasise the myriad of differences in the data and methodologies used when comparing this work to that of others.

During the graph optimisation process, the use of our custom similarity measure led to stronger GNN predictive performance than the cosine similarity method. This success can be attributed to the method's

normalisation by the maximum number of diagnoses between two patients. This normalisation accounts for the disparity in the overall number of diagnoses between patients. Patients with many diagnoses inherently have a higher chance of sharing some diagnoses with others simply due to the volume of diagnoses. By normalising by the maximum, the custom method mitigates this effect, ensuring that the similarity score does not disproportionately favour patients with more diagnoses. Unlike the cosine similarity, our custom method effectively emphasises overlaps in diagnoses and, by directly counting shared diagnoses, focuses on commonalities rather than the overall similarity of diagnosis patterns.

We integrated our custom approach with an optimisation process focused on the number of diagnoses and node linkages, restricting the latter to those with the highest similarity scores. By limiting the included diagnoses in this way, we ensure that only the most relevant diagnoses are retained, effectively reducing dimensionality. This reduction not only enhances the efficiency of GNN training but also improves the model's capacity for generalisation. Additionally, optimising node linkages ensures that only the most meaningful connections are preserved, maintaining critical relationships while minimising noise and enhancing the overall interpretability of the network.

It is important to note, however, that in our graph construction, edges were defined solely based on shared patient medical conditions (diagnoses), while other types of clinical data—such as patient vitals, medications, and laboratory tests—were reserved exclusively as node features. This design choice, while simplifying graph topology and reducing noise from potentially redundant connections, introduces a limitation by excluding potentially informative inter-patient similarities that could arise from non-diagnosis modalities. Alternative approaches, such as Patient-GAT [52], demonstrate how multimodal data fusion can be integrated directly into edge construction and attention mechanisms, enabling richer representations of complex patient relationships. Their method dynamically incorporates weighted similarities across heterogeneous data types, allowing for a more nuanced graph structure that may better reflect underlying clinical patterns. Future work could explore similar multimodal edge definitions to assess whether incorporating these additional relationships enhances predictive performance or interpretability.

GNN models, like any ML model, must be interpretable and trustworthy to be effectively utilised as decision-support tools in clinical settings. Understanding the factors (e.g., variables) that influence predictions is essential for validating the model's effectiveness and reliability through alignment with established medical knowledge. This validation process not only aids in model refinement but also ensures that the predictions are based on clinically relevant information. To this end, our feature importances and attention visualisations for the GNNs provide valuable insights, going as far as including patient-specific interpretations. Moreover, comparing these interpretations across different models—both traditional and advanced—can further enhance our understanding of their performance and applicability, thereby facilitating more informed and effective use in practice. Referring to this study as an example, we have found that the mean respiratory rate and patient history information play a significant factor in our GNN models. Furthermore, with methods to link patients, in this case with diagnoses, GNNs allow a further layer of depth to prediction tasks.

With regard to our focus on AF risk prediction, interpretative analyses have provided valuable insights that deepen our understanding of the underlying mechanisms. Specifically, the inclusion of patient clinical history has been shown to significantly enhance model performance by capturing longitudinal and contextual data that are crucial for accurate prediction. Moreover, the features most relevant to AF prediction appear to vary between individuals, underscoring the importance of personalised approaches in predictive modelling. Additionally, attention visualisations generated by the model have proven useful in uncovering feature-specific characteristics, allowing the identification of shared patterns and relationships among patients with similar clinical profiles. These patterns, revealed through the application of GNNs, offer critical insights for both researchers and the medical community, supporting the identification of key attributes that inform medical evaluations, improve triage processes, and ultimately guide more targeted interventions.

The proposed methodological approach represents a pivotal advancement towards achieving more accurate and meaningful representations of patient data, which serve as a foundation for the personalised management of AF-related stroke, a key objective of large-scale initiatives such as the European project TARGET [53,54]. By prioritising the most relevant diagnoses and connections, the model effectively captures patient-specific nuances, thereby facilitating tailored interventions and fostering improved clinical outcomes.

Finally, although this study prioritised the risk prediction of AF as an adverse event, the proposed methodology is highly versatile and could be readily adapted to predict a wide range of other adverse clinical events, offering broad applicability in healthcare research and clinical decision-making.

5. Conclusions

This study analysed the use of GNNs to develop a model that allows for the prediction of the risk (probability) of developing AF for critically ill stroke patients from information present in the MIMIC-IV database. Our models allow for the exploration of the multilateral associations of adverse risk, developing links between patients based on similar diagnoses and characteristics. As part of the GAT and GCN model construction, we assessed two methods of obtaining the similarity between patients; the typical cosine distance and our own novel approach that evaluates and compares diagnoses between patients. In this context, we found that GNN models present a better risk prediction performance compared to that of traditional Logistic Regression, XGBoost and Random Forest methods, obtaining an AUC score of 4 percent better.

CRediT authorship contribution statement

Cristian Y. Rivera-Juzga: Formal analysis, Data curation. **Dharmesh Mistry:** Writing – review & editing, Writing – original draft, Validation. **Adam T. Knowles:** Writing – review & editing, Writing – original draft, Validation. **Gregory Y.H. Lip:** Writing – review & editing, Validation, Funding acquisition. **Sandra Ortega-Martorell:** Writing – review & editing, Methodology, Funding acquisition, Formal analysis. **Ivan Olier:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Data availability statement

The MIMIC-IV database is available on the PhysioNet portal (https://physionet.org/content/mimiciv/2.2/) for credentialed users.

Ethics statement

This paper uses patient data from the MIMIC-IV database, available on the PhysioNet portal (https://physionet.org/content/mimiciv/2.2/) for credentialed users. It does not involve primary data collection.

The authors confirm that the work complies with the ethical standards for scholarly publishing, including:

- **Originality and Plagiarism**: The manuscript is original, has not been published elsewhere, and does not contain plagiarised material. All sources are appropriately cited and referenced.
- **Conflict of Interest**: There are no conflicts of interest that could influence the outcomes or interpretations presented in this paper.
- **Acknowledgment of Sources**: All studies, data, and other materials referenced in this paper are duly acknowledged to ensure transparency and credit to the original authors.

- **Compliance with Journal Guidelines**: The manuscript adheres to the ethical requirements and policies of the journal.

Funding

This research was partially funded by the EU project TARGET, which has received funding from the EU HORIZON EUROPE framework programme for research and innovation under Grant Agreement No. 101136244.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: **D.M. -** Team member of the TARGET project on health virtual twins for personalised management of atrial fibrillation and stroke (grant agreement no. 101136244).**A.T.K. -** Team member of the TARGET project on health virtual twins for personalised management of atrial fibrillation and stroke (grant agreement no. 101136244).**G.Y.H.L.** - Consultant and speaker for BMS/Pfizer, Boehringer Ingelheim, Daiichi-Sankyo, and Anthos. No fees are received personally. He is a National Institute for Health and Care Research (NIHR) Senior Investigator and co-PI of the AFFIRMO project on multimorbidity in AF (grant agreement no. 899871), TARGET project on health virtual twins for personalised management of atrial fibrillation and stroke (grant agreement no. 101136244) and ARISTOTELES project on artificial intelligence for the management of chronic long-term conditions (grant agreement no. 101080189), which are all funded by the EU's Horizon Europe Research & Innovation programme.**S.O.M.** - Principal Investigator of the TARGET project on health virtual twins for personalised management of atrial fibrillation and stroke (grant agreement no. 101136244) and senior investigator in the ARISTOTELES project on artificial intelligence for the management of chronic long-term conditions (grant agreement no. 101080189), both funded by the EU's Horizon Europe Research & Innovation programme. She is also a member of the board of the ART (Ageing Research Translation) of Healthy Ageing Network funded by the Biotechnology and Biological Sciences Research Council (BBSRC). No fees are received personally.**I.O.** - Methodological lead of the TARGET project on health virtual twins for personalised management of atrial fibrillation and stroke (grant agreement no. 101136244) and partner lead in the ARISTOTELES project on artificial intelligence for the management of chronic long-term conditions (grant agreement no. 101080189), both funded by the EU's Horizon Europe Research & Innovation programme. No fees are received personally.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2025.111095.

References

[1] N.A. Hilkens, B. Casolla, T.W. Leung, F.E. de Leeuw, Stroke, Lancet 403 (2024) 2820–2836, https://doi.org/10.1016/S0140-6736(24)00642-1.
[2] V.L. Feigin, M. Brainin, B. Norrving, S. Martins, R.L. Sacco, W. Hacke, M. Fisher, J. Pandian, P. Lindsay, World stroke organization (WSO): global stroke fact sheet 2022, Int. J. Stroke 17 (2022) 18–29, https://doi.org/10.1177/17474930211065917.
[3] L.W. Zhou, M.G. Lansberg, A. de Havenon, Rates and reasons for hospital readmission after acute ischemic stroke in a US population-based cohort, PLoS One 18 (2023) e0289640, https://doi.org/10.1371/journal.pone.0289640.
[4] B.J.R. Buckley, S.L. Harrison, A. Hill, P. Underhill, D.A. Lane, G.Y.H. Lip, Stroke-heart syndrome: incidence and clinical outcomes of cardiac complications following stroke, Stroke 53 (2022) 1759–1763, https://doi.org/10.1161/STROKEAHA.121.037316.
[5] K.L. Hoad, H. Jones, G. Miller, A.H. Abdul-Rahim, G.Y.H. Lip, B.J.R. Buckley, Stroke-heart syndrome: incidence and clinical outcomes of cardiac complications following intracerebral haemorrhage, Eur Stroke J (2024), https://doi.org/10.1177/23969873241264115.
[6] T.F. Chao, T.S. Potpara, G.Y.H. Lip, Atrial fibrillation: stroke prevention, Lancet Regional. Health - Europe 37 (2024) 100797, https://doi.org/10.1016/j.lanepe.2023.100797.
[7] G.A. Roth, G.A. Mensah, C.O. Johnson, et al., Global burden of cardiovascular diseases and risk factors, 1990-2019: update from the GBD 2019 study, J. Am. Coll. Cardiol. 76 (2020) 2982–3021, https://doi.org/10.1016/j.jacc.2020.11.010.
[8] L.A. Sposato, G.Y.H. Lip, K.G. Haeusler, Atrial fibrillation first detected after stroke: is timing and detection intensity relevant for stroke risk? Eur. Heart J. 45 (2024) https://doi.org/10.1093/eurheartj/ehad744.
[9] D.J. Gladstone, M. Spring, P. Dorian, V. Panzov, K.E. Thorpe, J. Hall, H. Vaid, M. O'Donnell, A. Laupacis, R. Côté, M. Sharma, J.A. Blakely, A. Shuaib, V. Hachinski, S.B. Coutts, D.J. Sahlas, P. Teal, S. Yip, J.D. Spence, B. Buck, S. Verreault, L.K. Casaubon, A. Penn, D. Selchen, A. Jin, D. Howse, M. Mehdiratta, K. Boyle, R. Aviv, M.K. Kapral, M. Mamdani, Atrial fibrillation in patients with cryptogenic stroke, N. Engl. J. Med. 370 (2014) 2467–2477, https://doi.org/10.1056/NEJMoa1311376.
[10] L.A. Sposato, S. Chaturvedi, C.-Y. Hsieh, C.A. Morillo, H. Kamel, Atrial fibrillation detected after stroke and transient ischemic attack: a novel clinical concept challenging current views, Stroke 53 (2022), https://doi.org/10.1161/STROKEAHA.121.034777.
[11] C. Dussault, H. Toeg, M. Nathan, Z.J. Wang, J.-F. Roux, E. Secemsky, Electrocardiographic monitoring for detecting atrial fibrillation after ischemic stroke or transient ischemic attack, Circ. Arrhythm. Electrophysiol. 8 (2015) 263–269, https://doi.org/10.1161/CIRCEP.114.002521.
[12] L.A. Sposato, L.E. Cipriano, G. Saposnik, E.R. Vargas, P.M. Riccio, V. Hachinski, Diagnosis of atrial fibrillation after stroke and transient ischaemic attack: a systematic review and meta-analysis, Lancet Neurol. 14 (2015) 377–387, https://doi.org/10.1016/S1474-4422(15)70027-X.
[13] J.P. Singh, J. Fontanarava, G. de Massé, T. Carbonati, J. Li, C. Henry, L. Fiorina, Short-term prediction of atrial fibrillation from ambulatory monitoring ECG using a deep neural network, European Heart J. Digital. Health 3 (2022) 208–217, https://doi.org/10.1093/ehjdh/ztac014.
[14] A. Papadopoulou, D. Harding, G. Slabaugh, E. Marouli, P. Deloukas, Prediction of atrial fibrillation and stroke using machine learning models in UK biobank, Heliyon 10 (2024) e28034, https://doi.org/10.1016/j.heliyon.2024.e28034.
[15] T. k Balaji, C.S.R. Annavarapu, A. Bablani, Machine learning algorithms for social media analysis: a survey, Comput. Sci. Rev. 40 (2021) 100395, https://doi.org/10.1016/j.cosrev.2021.100395.
[16] H.-C. Yi, Z.-H. You, D.-S. Huang, C.K. Kwoh, Graph representation learning in bioinformatics: trends, methods and applications, Briefings Bioinf. 23 (2022), https://doi.org/10.1093/bib/bbab340.
[17] K. Song, H. Park, J. Lee, A. Kim, J. Jung, COVID-19 infection inference with graph neural networks, Sci. Rep. 13 (2023) 11469, https://doi.org/10.1038/s41598-023-38314-3.
[18] W. Jiang, Graph-based deep learning for communication networks: a survey, Comput. Commun. 185 (2022) 40–54, https://doi.org/10.1016/j.comcom.2021.12.015.
[19] P. Pradhyumna, G.P. Shreya, Mohana, Graph Neural Network (GNN) in image and video understanding using deep learning for computer vision applications, in: 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), IEEE, 2021, pp. 1183–1189, https://doi.org/10.1109/ICESC51422.2021.9532631.
[20] Q. Zhang, J. Chang, G. Meng, S. Xu, S. Xiang, C. Pan, Learning graph structure via graph convolutional networks, Pattern Recogn. 95 (2019) 308–318, https://doi.org/10.1016/j.patcog.2019.06.012.
[21] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, P.S. Yu, A comprehensive survey on graph neural networks, IEEE Transact. Neural Networks Learn. Syst. 32 (2021) 4–24, https://doi.org/10.1109/TNNLS.2020.2978386.
[22] B. Khemani, S. Patil, K. Kotecha, S. Tanwar, A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions, J. Big Data 11 (2024) 18, https://doi.org/10.1186/s40537-023-00876-4.
[23] P. Veličković, A. Casanova, P. Liò, G. Cucurull, A. Romero, Y. Bengio, Graph attention networks, in: 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings, 2018, https://doi.org/10.1007/978-3-031-01587-8_7.
[24] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, M. Sun, Graph neural networks: a review of methods and applications, AI Open 1 (2020) 57–81, https://doi.org/10.1016/j.aiopen.2021.01.001.
[25] S. Georgousis, M.P. Kenning, X. Xie, Graph deep learning: state of the art and challenges, IEEE Access 9 (2021) 22106–22140, https://doi.org/10.1109/ACCESS.2021.3055280.
[26] C. Tong, E. Rocheteau, P. Veličković, N. Lane, P. Liò, in: Predicting Patient Outcomes with Graph Representation Learning, 2022, pp. 281–293, https://doi.org/10.1007/978-3-030-93080-6_20.
[27] H. Oss Boll, A. Amirahmadi, M.M. Ghazani, W.O. de Morais, E.P. de Freitas, A. Soliman, F. Etminani, S. Byttner, M. Recamonde-Mendoza, Graph neural networks for clinical risk prediction based on electronic health records: a survey, J. Biomed. Inf. 151 (2024) 104616, https://doi.org/10.1016/j.jbi.2024.104616.
[28] Z. Xu, J. Guo, L. Qin, Y. Xie, Y. Xiao, X. Lin, Q. Li, X. Li, Predicting ICU interventions: a transparent decision support model based on multivariate time series graph convolutional neural network, IEEE J. Biomed. Health. Inform. 28 (2024) 3709–3720, https://doi.org/10.1109/JBHI.2024.3379998.
[29] A.E.W. Johnson, L. Bulgarelli, L. Shen, A. Gayles, A. Shammout, S. Horng, T. J. Pollard, S. Hao, B. Moody, B. Gow, L.H. Lehman, L.A. Celi, R.G. Mark, MIMIC-IV,

a freely accessible electronic health record dataset, Sci. Data 10 (2023) 1, https://doi.org/10.1038/s41597-022-01899-x.
[30] T.N. Kipf, M. Welling, Semi-Supervised Classification with Graph Convolutional Networks, 2016.
[31] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna: a next-generation hyperparameter optimization framework, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2019, https://doi.org/10.1145/3292500.3330701.
[32] M. Scott, Lundberg and Su-In lee, A unified approach to interpreting model predictions, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, pp. 4768–4777.
[33] M. Fey, J.E. Lenssen, Fast graph representation learning with PyTorch geometric, in: ICLR Workshop on Representation Learning on Graphs and Manifolds, 2019.
[34] L. van der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (2008) 2579–2605.
[35] G. Teasdale, B. Jennett, Assessment of coma and impaired consciousness, Lancet 304 (1974) 81–84, https://doi.org/10.1016/S0140-6736(74)91639-0.
[36] S. Tetri, L. Mäntymäki, S. Juvela, P. Saloheimo, J. Pyhtinen, H. Rusanen, M. Hillbom, Impact of ischemic heart disease and atrial fibrillation on survival after spontaneous intracerebral hemorrhage, J. Neurosurg. 108 (2008) 1172–1177, https://doi.org/10.3171/JNS/2008/108/6/1172.
[37] B. Morseth, B. Geelhoed, A. Linneberg, L. Johansson, K. Kuulasmaa, V. Salomaa, L. Iacoviello, S. Costanzo, S. Söderberg, T.J. Niiranen, J.K.K. Vishram-Nielsen, I. Njølstad, T. Wilsgaard, E.B. Mathiesen, M.-L. Løchen, T. Zeller, S. Blankenberg, F. M. Ojeda, R.B. Schnabel, Age-specific atrial fibrillation incidence, attributable risk factors and risk of stroke and mortality: results from the MORGAM consortium, Open Heart 8 (2021) e001624, https://doi.org/10.1136/openhrt-2021-001624.
[38] D. Aune, Y. Mahamat-Saleh, E. Kobeissi, T. Feng, A.K. Heath, I. Janszky, Blood pressure, hypertension and the risk of atrial fibrillation: a systematic review and meta-analysis of cohort studies, Eur. J. Epidemiol. 38 (2023) 145–178, https://doi.org/10.1007/s10654-022-00914-0.
[39] R. Ying, D. Bourgeois, J. You, M. Zitnik, J. Leskovec, GNNExplainer: generating explanations for graph neural networks, in: Proceedings of Neurips, 2019, pp. 9244–9255.
[40] C.B. Christiansen, T.A. Gerds, J.B. Olesen, S.L. Kristensen, M. Lamberts, G.Y.H. Lip, G.H. Gislason, L. Køber, C. Torp-Pedersen, Atrial fibrillation and risk of stroke: a nationwide cohort study, Europace 18 (2016) 1689–1697, https://doi.org/10.1093/europace/euv401.
[41] A. Alshehri, Stroke in atrial fibrillation: review of risk stratification and preventive therapy, J. Family Commun. Med. 26 (2019) 92, https://doi.org/10.4103/jfcm.JFCM_99_18.
[42] S. Elsheikh, A. Hill, G. Irving, G.Y.H. Lip, A.H. Abdul-Rahim, Atrial fibrillation and stroke: state-Of-the-art and future directions, Curr. Probl. Cardiol. 49 (2024) 102181, https://doi.org/10.1016/j.cpcardiol.2023.102181.
[43] R.R. Huxley, K.B. Filion, S. Konety, A. Alonso, Meta-analysis of cohort and case–control studies of type 2 diabetes mellitus and risk of atrial fibrillation, Am. J. Cardiol. 108 (2011) 56–62, https://doi.org/10.1016/j.amjcard.2011.03.004.
[44] L.J. Bohne, D. Johnson, R.A. Rose, S.B. Wilton, A.M. Gillis, The association between diabetes mellitus and atrial fibrillation: clinical and mechanistic insights, Front. Physiol. 10 (2019), https://doi.org/10.3389/fphys.2019.00135.
[45] A. Alonso, S.K. Agarwal, E.Z. Soliman, M. Ambrose, A.M. Chamberlain, R. J. Prineas, A.R. Folsom, Incidence of atrial fibrillation in whites and african-Americans: the atherosclerosis risk in communities (ARIC) study, Am. Heart J. 158 (2009) 111–117, https://doi.org/10.1016/j.ahj.2009.05.010.
[46] R.M.J. van der Velden, A.N.L. Hermans, N.A.H.A. Pluymaekers, M. Gawalko, A. Elliott, J.M. Hendriks, F.M.E. Franssen, A.M. Slats, V.P.M. van Empel, I.C. Van Gelder, D.H.J. Thijssen, T.M.H. Eijsvogels, C. Leue, H.J.G.M. Crijns, D. Linz, S. O. Simons, Dyspnea in patients with atrial fibrillation: mechanisms, assessment and an interdisciplinary and integrated care approach, Int. J. Cardiol. Heart Vasc. 42 (2022) 101086, https://doi.org/10.1016/j.ijcha.2022.101086.
[47] S. Ortega-Martorell, I. Olier, M. Ohlsson, G.Y.H. Lip, Advancing personalised care in atrial fibrillation and stroke: the potential impact of AI from prevention to rehabilitation, Trends Cardiovasc. Med. (2024), https://doi.org/10.1016/J.TCM.2024.12.003.
[48] I. Olier, S. Ortega-Martorell, M. Pieroni, G.Y.H. Lip, How machine learning is impacting research in atrial fibrillation: implications for risk prediction and future management, Cardiovasc. Res. 117 (2021) 1700–1717, https://doi.org/10.1093/cvr/cvab169.
[49] X. Xu, S. Wei, C. Ma, K. Luo, L. Zhang, C. Liu, Atrial fibrillation beat identification using the combination of modified frequency slice wavelet transform and convolutional neural networks, J. Healthc. Eng. 2018 (2018) 1–8, https://doi.org/10.1155/2018/2102918.
[50] O. Faust, A. Shenfield, M. Kareem, T.R. San, H. Fujita, U.R. Acharya, Automated detection of atrial fibrillation using long short-term memory network with RR interval signals, Comput. Biol. Med. 102 (2018) 327–335, https://doi.org/10.1016/j.compbiomed.2018.07.001.
[51] S. Ortega-Martorell, M. Pieroni, B.W. Johnston, I. Olier, I.D. Welters, Development of a risk prediction model for new episodes of atrial fibrillation in medical-surgical critically ill patients using the AmsterdamUMCdb, Front. Cardiovasc. Med. 9 (2022), https://doi.org/10.3389/fcvm.2022.897709.
[52] C. Xiao, N. Pham, E. Imel, X. Luo, Patient-GAT: sarcopenia prediction using multi-modal data fusion and weighted graph attention networks, in: Proceedings of the ACM Symposium on Applied Computing, 2023, https://doi.org/10.1145/3555776.3578731.
[53] S. Ortega-Martorell, I. Olier, M. Ohlsson, G.Y.H. Lip, TARGET: a major European project aiming to advance the personalised management of atrial fibrillation related stroke, Thromb. Haemost. (2024), https://doi.org/10.1055/a-2438-5671.
[54] S. Ortega-Martorell, I. Olier, G.Y.H. Lip, TARGET consortium, A European network to develop virtual twin technology for personalized stroke management in atrial fibrillation: the TARGET consortium, Eur. Heart J. (2024), https://doi.org/10.1093/EURHEARTJ/EHAE673.