Network Modeling Analysis in Health Informatics and Bioinformatics (2026) 15:9
https://doi.org/10.1007/s13721-025-00690-3

ORIGINAL ARTICLE

A feasible machine learning framework for diagnosing stroke patients versus mimic conditions incorporating extracellular vesicle characterization and EHR features fusion

Vasileios E. Katsigiannis[1,2] · Ioannis Kakkos[1,2] · Stavros Theofanis Miloulis[1,2] · Ioannis A. Vezakis[1,2] · Ourania Petropoulou[1,2] · Joyce Rops[3,4,5] · Naomi C. Buntsma[3,4,5] · Edwin van der Pol[5,3] · Dimitrios I. Fotiadis[6] · George K. Matsopoulos[1,2]

Received: 21 August 2025 / Revised: 24 November 2025 / Accepted: 26 November 2025

Abstract

Stroke remains a leading cause of morbidity and mortality worldwide, necessitating rapid and accurate diagnosis to improve patient outcomes. The gold standard in stroke diagnosis relies on a costly and time-consuming combination of clinical evaluation and neuroimaging techniques, while also being associated with misdiagnoses due to mimicking conditions. This highlights the need for improving the current healthcare system towards faster, more accurate, and low-resource diagnostics. In this direction, this study presents the development of a novel artificial intelligence-driven decision support system for stroke diagnosis using machine learning techniques to fuse standard EHR variables (electronic health records) with information derived from extracellular vesicles (EVs) in blood plasma, which hold promise as diagnostic markers due to their association with stroke pathophysiology. As such, flow cytometry was employed to measure blood samples from 140 patients suffering from stroke or stroke-mimicking conditions. The derived EV-features, including statistical characteristics of their diameters' distribution and concentration, were integrated with demographical, clinical, and biochemical patient data to form a comprehensive feature set. By utilizing a decision tree learning model, the proposed system achieved a classification weighted accuracy of 86%. This achievement indicates the approach's potential for integrating cutting-edge photonics-based signals with standard HER information to improve clinical decision-making in emergency settings and for aiding treatment strategies in stroke care.

Keywords Stroke · Machine learning · Biomarkers · Extracellular vesicles · Decision support system

✉ Vasileios E. Katsigiannis
vassiliskat@biomed.ntua.gr

[1] Biomedical Engineering Laboratory, National Technical University of Athens, 9, Iroon Polytechniou Str, Zografos, 15773 Athens, Greece

[2] Biomedical Engineering Laboratory, Institute of Communication and Computer Systems, Patission 42, 10682 Athens, Greece

[3] Laboratory of Experimental Clinical Chemistry and Amsterdam Vesicle Center, Laboratory Specialized Diagnostics & Research, Department of Laboratory Medicine, Amsterdam UMC, University of Amsterdam, Meibergdreef 9, Amsterdam, Netherlands

[4] Department of Neurology, Amsterdam UMC, University of Amsterdam, Meibergdreef 9, Amsterdam, Netherlands

[5] Biomedical Engineering & Physics, Amsterdam UMC, University of Amsterdam, Meibergdreef 9, Amsterdam, Netherlands

[6] Unit of Medical Technology and Intelligent Information Systems, Dept. of Materials Science and Engineering, University of Ioannina, Ioannina, Greece

Published online: 19 December 2025

1 Introduction

Stroke is defined as a sudden disruption of brain function resulting from an interruption in cerebral blood flow. It affects more than 15 million individuals worldwide each year, with survivors often facing long-term complications like motor, cognitive, and speech impairments (Benjamin et al. 2018). Stroke may be classified into two major types: ischemic stroke (which occurs when a thrombus or embolus obstructs a cerebral artery, causing rapid neuronal death), and hemorrhagic stroke (which arises from the rupture of a blood vessel, causing intracranial bleeding, increased pressure, and subsequent tissue damage) (Leppälä et al. 1999; Wilson et al. 2018). A closely related condition is the Transient Ischemic Attack (TIA), often referred to as a "mini-stroke", where blood flow to the brain is temporarily reduced, leading to stroke-like symptoms that resolve within 24 h (Shah and Edlow 2004). Timely intervention is essential with an optimal therapeutic window for ischemic stroke (intravenous thrombolysis with tissue plasminogen activator) being 4.5 h from symptom onset (Widimsky et al. 2023). Similarly, hemorrhagic stroke necessitates rapid treatment to control bleeding and reduce intracranial pressure. However, stroke and TIA symptoms may overlap with other neurological conditions (such as epileptic seizures or primary headache disorders) which can lead to misdiagnosis and inappropriate management, potentially exacerbating patient outcomes (Yan et al. 2016).

Currently, stroke diagnosis currently relies on a combination of clinical assessment and neuroimaging techniques with Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) primarily employed to differentiate ischemic from hemorrhagic strokes (Jadhav et al. 2020). Additionally, CT-Angiography and MR-Angiography enable the detection of large vessel occlusions and intracranial bleeding (Patil et al. 2022). Although these imaging techniques provide high diagnostic accuracy, they require infrastructure access, they can be expensive and may (in the case of CT) involve ionizing radiation exposure (Meyers et al. 2011; Martinez-Gutierrez et al. 2023). More importantly, since neuroimaging procedures can be time-consuming and treatment efficacy is highly dependent on prompt intervention, achieving a rapid and accurate diagnosis is essential for lowering the rate of misdiagnoses which has been reported in 1.2%–12.7% of cases (Candelaresi et al. 2023). In this context, ongoing research is increasingly focused on identifying novel biomarkers, such as proteins, RNAs, lipids, metabolites, and extracellular vesicles (EVs) (Kim et al. 2013; Kamtchum-Tatuene and Jickling 2019). As such, EVs are nanoscale, membrane-enclosed particles released by all types of cells into their surrounding environment. Consequently, bodily fluids, such as blood plasma, saliva, amniotic fluid, and urine, contain EVs. Their composition and concentration vary in response to physiological and pathological states, making EVs promising biomarkers of disease (Yuana et al. 2013; Welsh et al. 2023). Such information can be extracted from biological samples using techniques based on photonic principles such as flow cytometry (FCM). In the case of stroke, EVs originating from neurons and endothelial cells carry molecular signatures indicative of brain injury, while actively participating in stroke-associated inflammation, endothelial dysfunction, and neurodegenerative processes (Yang et al. 2018). Moreover, stroke patients exhibit distinct EV profiles in their blood, suggesting that EVs could serve as non-invasive biomarkers, facilitating timely and accurate diagnosis (Couch et al. 2017; Stenz et al. 2020a). However, despite their promising potential, current EV findings remain heterogeneous and sometimes contradictory. Several studies have reported elevations in EV concentrations and distinct molecular profiles in acute ischemic or hemorrhagic stroke patients compared to controls (Kim et al. 2025; Reymond et al. 2025), whereas others found the reverse trend or no consistent association between EV levels and stroke subtype, severity, or clinical outcomes (Casado-Fernández et al. 2024; Buntsma et al. 2025).

Nevertheless, new technologies such as artificial intelligence (AI) can analyze high-dimensional and complex medical data, identifying subtle patterns not discernible to clinicians, supporting clinical decision-making, and reducing the likelihood diagnostic errors (Bates et al. 2014; Topol 2019). Regarding stroke-diagnosis, several studies applied AI-based approaches to support and improve conventional diagnostic procedures (Czap et al. 2022). These systems have demonstrated accuracy levels comparable to those of clinical experts, indicating that AI has the potential to substantially increase both the speed and precision of stroke identification. For instance, Heo et al. developed a machine learning (ML) model utilizing clinical variables, (such as age, blood pressure, and diabetes status), to predict ischemic stroke outcomes (Heo et al. 2019). Their model outperformed conventional statistical methods, highlighting ML's potential to integrate multiple data types and offer more accurate predictions for stroke patients. Moreover, the integration of AI with biomarker-based diagnostics represents an emerging and promising direction, detecting and identifying signatures indicative of stroke undetectable using standard diagnostic tools (Mainali et al. 2021). In this context, Liu et al. employed ML algorithms to analyze a panel of blood-based biomarkers, demographics, and dietary features, achieving over 75% classification accuracy (Liu et al. 2022). In a similar way, Gkantzios et al. utilized ML to evaluate blood biomarkers and parameters for the prediction of stroke survivors' functional outcomes, suggesting that C-reactive protein (CRP) and Erythrocyte

Sedimentation Rate levels may be important predictors of functional outcomes in stroke patients (Gkantzios et al. 2023). Even though blood panels show promising accuracy for acute stroke classification but face assay turnaround and standardization issues, highlighting the need of rapid assays and compact, robust panels for triage (Rahmig et al. 2024).

However, there remains a wide range of unexplored features that could further deepen our understanding of stroke pathophysiology. In this context, cell-state information relevant to stroke pathophysiology which is encoded by profiling the EVs can be combined with routine EHR variables may provide a fast, low-resource triage, supporting early recognition, streamlining clinical workflows, and assisting healthcare professionals in making data-driven decision (Théry et al. 2018). Building on this rationale, the present study aims: (1) to develop an AI-based Decision Support System (DSS) for the automated diagnosis of stroke occurrences, involving the use of EVs as novel stroke biomarkers, while incorporating clinical and biochemical parameters; (2) to enhance interpretability of the underlying pathophysiological mechanisms contributing to stroke.

In this study, we present a framework that integrates EV-derived features from calibrated flow cytometry with routinely collected EHR variables to differentiate stroke from stroke-mimicking conditions. As such, we quantify EV size-distribution statistics and concentration and fuse these measurements with biochemical and demographic EHR characteristics feeding them to a feature selection and classification procedure to provide efficient discrimination between stroke and stroke mimic conditions. Our approach employes the ReliefF feature selection method along with several ML classifiers to identify a small number of robust predictors, ensuring high classification performance and clinical interpretability. To the best of our knowledge, this is the first investigation to utilize particle EV flow-cytometry profiles together with EHR-derived clinical variables in a unified machine learning framework for accurate stroke versus stroke-mimic classification in the clinical settings. The organization of this paper is as follows: Sect. 2 describes the study population, EV acquisition pipeline, feature extraction procedure, and machine learning methodology; Sect. 3 presents the classification results and feature importance findings; Sect. 4 discusses the implications, limitations, and biological interpretation of the results; and Sect. 5 outlines lessons learned and future research directions.

Table 1 Demographical and clinical characteristics of the study

Stroke		Stroke mimic	
Number of Individuals	107	Number of Individuals	33
Female, N (%)	55 (51.4)	Female, N (%)	15 (45.5)
Age, mean in years (std, range)	70 (12.3, 31–98)	Age, mean in years (std, range)	61 (12.7, 31–83)

2 Materials and methods

2.1 Subjects

The dataset comprised 140 participants (males/females = 70/70, mean age = 64.8 ± 12 years/70.5 ± 13 years), divided into two groups based on stroke occurrence. The clinical and demographic characteristics of all participants were evaluated to determine their eligibility for inclusion in this study (Table 1). The control group consisted of 33 individuals with stroke-mimicking conditions, including primary headaches ($n = 12$), epilepsy ($n = 13$), and vestibular disorders ($n = 8$). The stroke group included 107 patients diagnosed with hemorrhagic stroke ($n = 16$), ischemic stroke ($n = 62$), or transient ischemic attack ($n = 24$). Prior to all processing and analysis procedures, data were fully anonymized.

2.2 Data acquisition

The data used in this study were collected as part of an observational clinical investigation conducted at the Amsterdam University Medical Center within the framework of the CINTICS II project. In detail, blood samples were obtained from 481 patients over 18 years old who admitted to the AMC Emergency Department with suspected stroke. From this cohort, 159 patients were selected for preliminary analysis based on the following criteria: (1) availability of >4 aliquots of blood plasma, (2) limited experimental capacity, (3) power calculations to achieve the primary endpoint, (4) age and sex balance among stroke subtypes and mimics. To maintain data integrity and analytical reliability, 19 patients with missing or incomplete data were subsequently excluded, resulting in a final dataset of 140 patient samples. The study protocol was reviewed and approved by the Ethical Review Board of Amsterdam University Medical Center, under a deferred consent procedure (approval number: NL72929.018.20).

2.2.1 Blood collection and plasma Preparation

Blood collection was performed through either antecubital venipuncture or from an intravenous drip (4.0 mL K2EDTA, 7.2 mg; BD Vacutainer®). The obtained blood samples were transported from the emergency room to the laboratory within 2 h after collection. Blood plasma was prepared by centrifuging whole blood at 2,500 g for 15 min at 20 °C, acceleration speed 9 and deceleration speed 1 (Rotina

380 H, Hettich Zentrifugen, Tuttlingen, Germany) (Lacroix et al. 2012, 2013). Plasma was collected by aspiration until 10 mm above the buffy coat and transferred to a new 12 mL tube (Greiner Bio-One GmbH, Kremsmünster, Austria). The collected plasma was centrifuged and collected again using the same method as used for whole blood. The collected plasma was transferred to 1.5 mL Eppendorf tubes (Thermo Fischer Scientific, Waltham, MA, United States). After vortexing, aliquots were stored at −80 °C until analysis.

2.2.2 Flow cytometry measurements

EV concentrations in the samples were determined using flow cytometry (Northern Lights, Cytek Biosciences, Fremont, CA, United States) (Fig. 1). To measure the concentration of EVs originating from cells of interest, EVs were immunofluorescently stained using labels specific for EVs binding lactadherin (Lact) conjugated to fluorescein (FITC). For the present study, the lactadherin-based (Lact-FITC+) measurements were incorporated since studies have shown that they excel at detecting stroke-related EVs because lactadherin binds with high affinity to the phosphatidylserine lipid, which becomes externalized on the surface of EVs during stroke occurring (Thulin et al. 2020; Li et al. 2024). The labeling reagents were centrifuged at 18,890 g for 5 min at 20 °C to remove aggregates. The supernatant minus 10 µl of the starting volume was collected and used for staining. Before staining, samples were individually diluted in Dulbecco's phosphate-buffered saline (DPBS) to ensure the count rate was below 30,000 events/s, which prevents swarm detection for plasma samples measured on our instrument (Buntsma et al. 2023). To stain EVs, 2.5 µl of the labelling reagent mix was incubated with 20 µl of the diluted sample for 2 h at room temperature in the dark. After incubation, 200 µl of DPBS was added to decrease background fluorescence. Stained samples were measured with side scatter triggering until either a volume of 150 µl or 10,000,000 counts was reached. On each measurement day, buffer-only controls (30 µl DPBS), positive controls (30 µl stained pooled plasma), and unstained controls were measured. Label dilutions were made weekly, after which the buffer with reagents control was measured.

EV diameters in nm were determined using Rosetta Calibration (v2.04, Exometry, Amsterdam, The Netherlands) based on side scattering signals and assuming a core refractive index of 1.38, a shell refractive index of 1.48, and a shell thickness of 6 nm. Fluorescence signals were calibrated and expressed as units of molecules of equivalent soluble fluorochromes (MESF) using custom-built software (MATLAB R2020b, MathWorks, Natick, MA, United States). Finally, the MESF-calibrated measurements were gated to effectively distinguish EV particles of diameters between 100 and 1000 nm from other entities in blood plasma. MIFlowCyt-EV and datafiles of the flow cytometry experiment are available via Figshare (see: https://doi.org/10.6084/m9.figshare.28688051).

2.3 Feature extraction

To enable AI-based analysis, EV-related data were transformed into a structured feature set. The EV diameter distribution was discretized into 100 nm-wide bins, from which statistical descriptors were calculated, including mean diameter, median diameter, kurtosis, skewness, entropy, and standard deviation (Std) of the distribution. Additional variables included the Lact-FITC$^+$ EV concentration, the binned Lact-FITC$^+$ EV concentration, and the total particle concentration. Figure 2 represents the gating of MESF-calibrated measurements to effectively distinguish the Lact-FITC$^+$ EV particles of diameters between 100 and 1000 nm from other

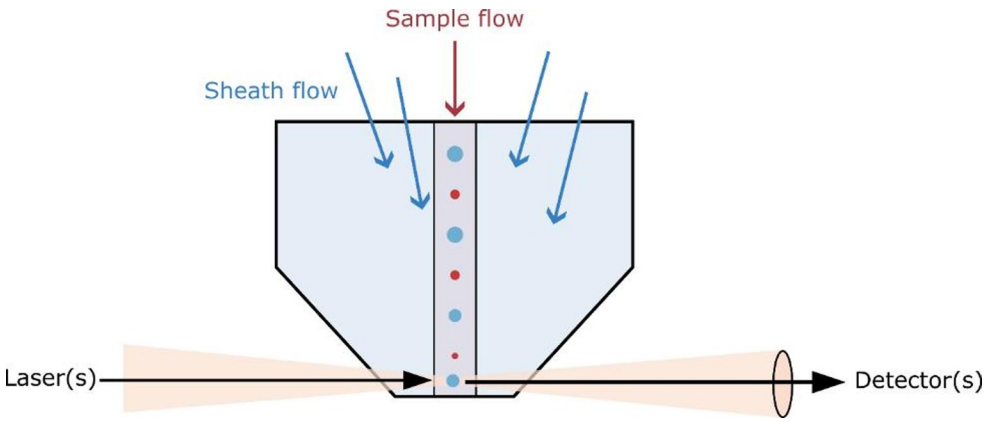

Fig. 1 Flow cytometry setup representation. Particles are hydrodynamically focused by sheath fluid into a laser beam, where their fluorescence and light-scatter signals are detected, electronically processed, and recorded. Accurate extracellular vesicle cytometry requires adherence to community reporting standards and control of swarm detection to avoid overcounting. Therefore, the calibration, the report instrument settings and the gating documentation must be in line with the MIFlowCyt-EV guidance to that mitigate swarm artifacts

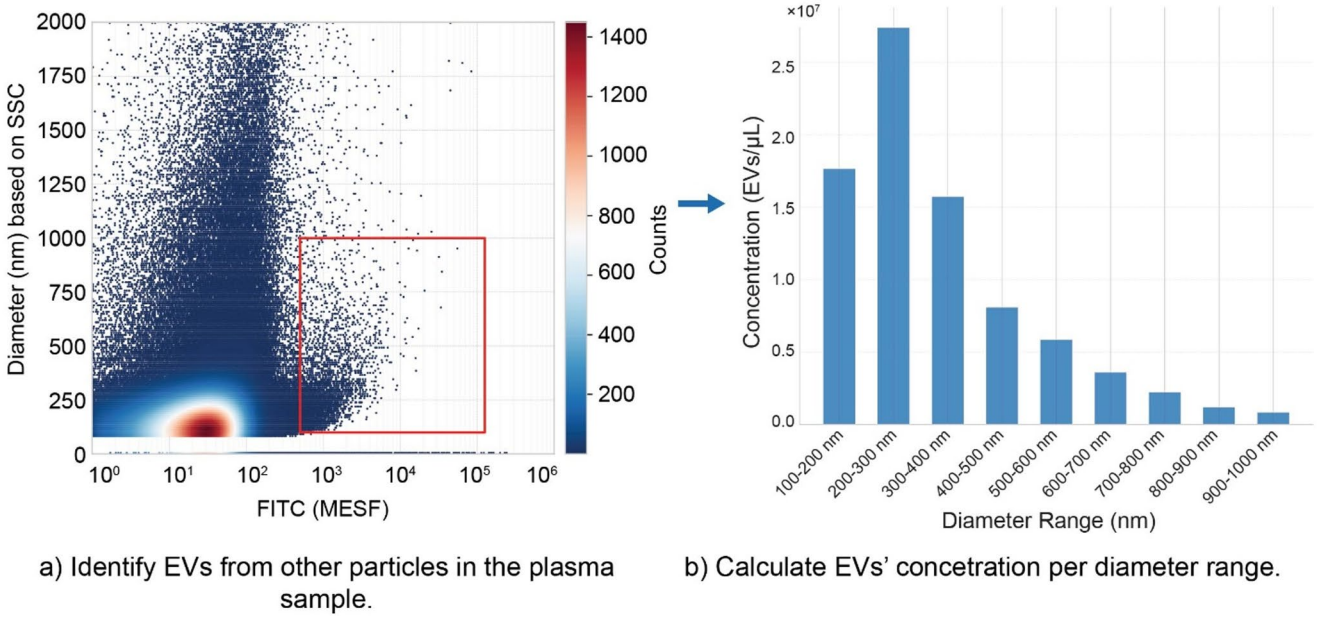

Fig. 2 Gating procedure implemented on MESF-calibrated measurements to accurately separate Lact-FITC$^+$ EVs in the 100–1000 nm size interval from other plasma-derived entities. In (**a**) the MESF-calibrated Flowcytometry measurements are displayed, as well as the gating limits of the Lact-FITC$^+$ EVs to distinguish among other blood particles. Each dot represents a particle measured from the Flow cytometer; in (**b**) the diameter distributions of Lact-FITC$^+$ EV particles are represented

Table 2 Summary of the included features

Type	Feature Names
EV-related	EV diameter distribution mean, median, kurtosis, skewness, entropy, standard deviation, bin concentration (9x), Total EV number, total EV concentration,
Demographic	Sex, age
Physiological metrics	RR syst, RR diast
Biochemical metrics	Thrombocytes

entities in plasma sample. These 17 EV-derived features were then combined with patient information and clinical parameters, including sex, age (in years), respiratory rate systolic (RR syst.), respiratory rate diastolic (RR diast.), and thrombocyte count, resulting in a total of 21 input features (Table 2). The statistical distributions of the electronic health record (EHR)-derived variables are illustrated in Fig. 3.

2.4 Feature selection and classification

In the first stage, only EV-specific features were incorporated to assess whether EV profiles alone could differentiate between the two diagnostic classes (i.e., stroke vs. stroke-mimics). As such, to reduce the dataset's dimensionality and retain only the most important features, the ReliefF feature selection method was applied. ReliefF identifies relevant features based on how well they distinguish between dataset instances that are near each other but belong to different classes (Robnik-Šikonja and Kononenko 2003). In detail, the ReliefF is a distance-based algorithm that iteratively samples instances from the dataset and evaluates each feature's ability to discriminate between similar instances of opposing classes. For each sampled instance, the method compares its values with those of its k-nearest neighbors from both the same and different classes, updating feature weights according to the degree of the observed differences. Features that consistently contribute to greater inter-class separation receive higher weights, thereby being considered more relevant for the classification task, as illustrated in Fig. 4. ReliefF was incorporated due to its ability to handle heterogeneous, possibly non-linear interactions and ranks features via near-neighbor class separability, and it is considered a robust filter for small and medium size datasets (Urbanowicz et al. 2018).

The available data were partitioned into training and test sets using a 70:30 split, stratified such that the proportion of patients with stroke and stroke mimics was preserved in both subsets. To mitigate potential class imbalance in the training set, instances of the minority class were randomly oversampled by duplication. All continuous predictors were then linearly rescaled to the [0,1] interval to place variables on a comparable scale and to facilitate stable optimization across the different classification algorithms.

Based on the rationale that EVs' characteristics, such as their concentration and size distribution, can provide insights into the underlying biological processes associated with stroke, widely used ML algorithms were employed to evaluate whether the selected features could accurately

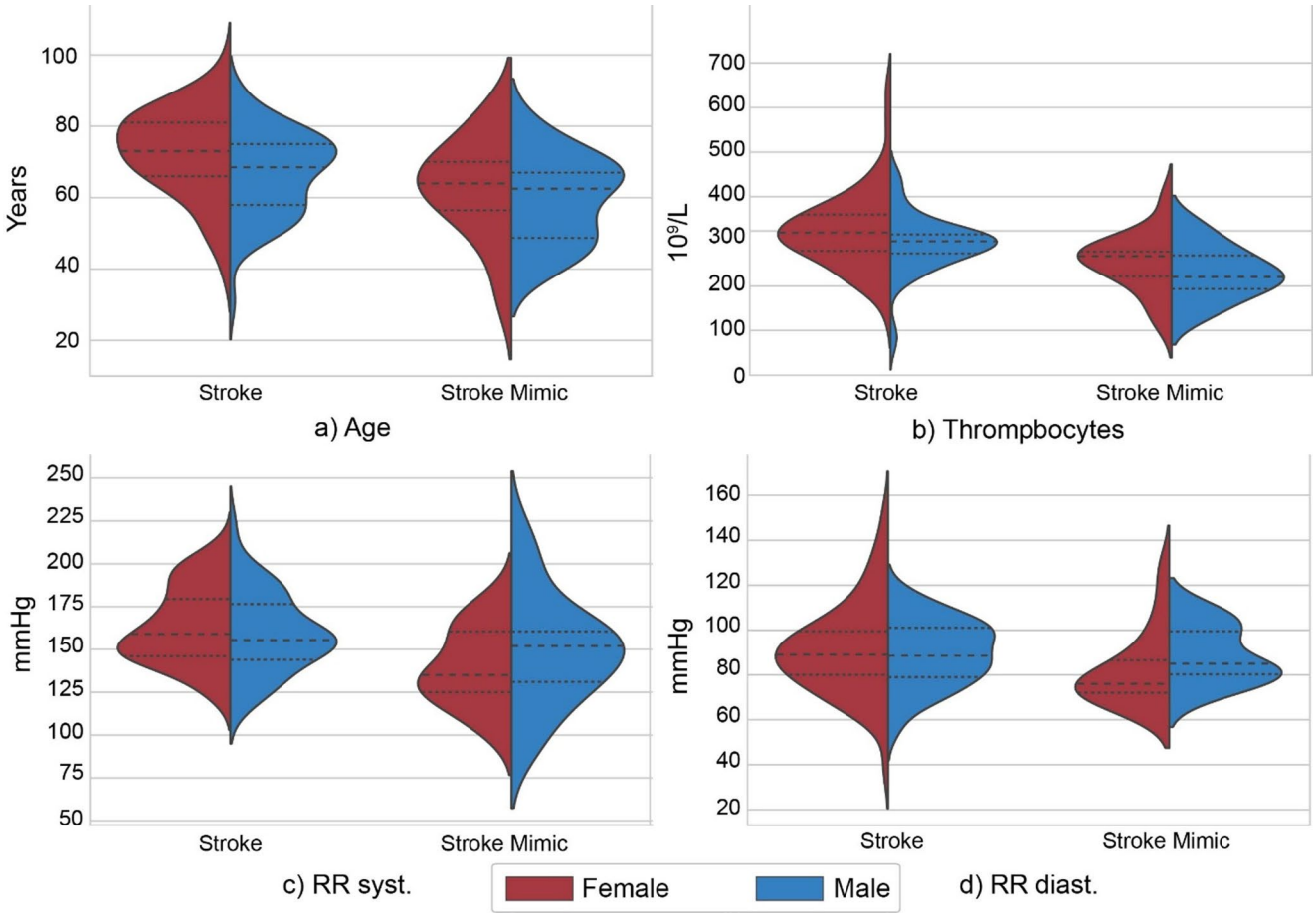

Fig. 3 Statistical distributions of demographic, physiological, and biochemical metrics for male and female populations among stroke and stroke mimic conditions

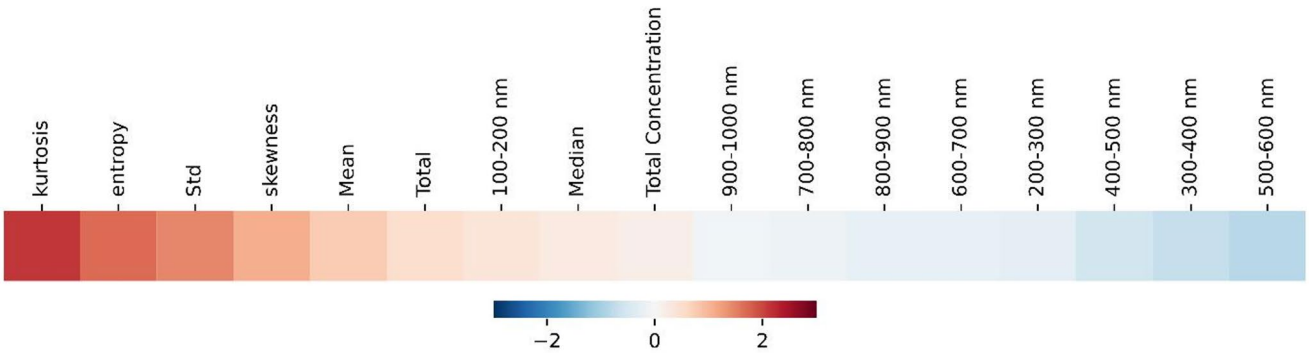

Fig. 4 Extracellular vesicles-related features score and ranking using the ReliefF feature selection method

classify patients as having either stroke or stroke-mimics in a classifier-agnostic manner, that is, whether consistently good performance could be achieved across multiple model families rather than relying on a single algorithm. Specifically, the classifiers applied included: Logistic Regression (LR) (Bisong 2019), Support Vector Machine (SVM) (Cervantes et al. 2020), Random Forest (RF) (Shaik and Srinivasan 2019), Stochastic Gradient Descent (SGD) (Liu et al. 2018), and Decision Tree (DT) (Charbuty and Abdulazeez 2021), so our assessment covers linear, margin-based, ensemble, online linear, and rule-based learners. Given the high-stakes due to their clinical use and modest sample size, we prioritized inherently interpretable models, the LR algorithm serves as a strong and interpretable linear baseline for binary classification tasks (estimating the log-odds of stroke as a linear combination of predictors), and thus allowing

direct inspection of the direction and relative magnitude of associations. Its probabilistic output and well-understood regularization properties make LR a standard reference method in clinical prediction modeling, particularly suitable when sample sizes are modest and interpretability is essential (Bisong 2019). Furthermore, SVMs are margin-based classifiers that maximize the separation between classes and are effective in high-dimensional settings. Linear and radial basis function (RBF) kernels are employed to capture, respectively, linear decision boundaries and moderate non-linear relationships between EV-derived and clinical variables. They have a strong theoretical foundation and documented performance in biomedical classification tasks, justifying their role as a robust non-probabilistic benchmark (Cervantes et al. 2020). In addition, RF is an ensemble of decision trees trained on bootstrapped samples and random feature subsets. It captures complex non-linearities and interactions between EV-derived and clinical variables, is relatively robust to outliers and noise, and provides measures of variable importance (Shaik and Srinivasan 2019). Similarly, decision trees recursively partition the predictor space into axis-aligned decision rules and produce explicit, human-readable decision paths. They inherently model interactions between variables, such as combinations of EV distributional features with blood pressure or age, and capture non-linear relationships without requiring prior feature transformation (Charbuty and Abdulazeez 2021). Moreover, the SGD-based classifier implements regularized linear models optimized via incremental gradient updates. It offers a computationally efficient way to fit penalized linear decision functions and serves as an additional linear benchmark with strong regularization control (Liu et al. 2018).

The classification performance was evaluated using a set of typical classification metrics, including precision (indicating the probability that a case predicted as stroke truly represents a stroke event and thus the reliability of positive alerts); recall (quantifying the proportion of true strokes correctly identified); F1-score (calculated as the harmonic mean of precision and recall, to summarize performance when both false negatives and false positives are important); the area under the receiver operating characteristic curve (ROC-AUC) (as an independent measure of discriminative ability); accuracy and weighted accuracy, which accounts for class imbalances and is defined as the average of the accuracies of the stroke and stroke-mimic classes and will be referenced as accuracy in the context of the present study (Vujovic 2021).

Model development and selection were conducted exclusively on the training set. Within this subset, 5-fold cross-validation was employed to enhance generalizability to unseen data and to tune model hyperparameters. Specifically, the training data were randomly partitioned into five approximately equal folds; in each iteration, four folds were used to train the model and the remaining fold served as the validation set. This process was repeated so that each fold acted once as the validation set, and performance estimates were averaged across folds, providing a robust assessment of out-of-sample behavior and reducing the risk of overfitting (Rudin 2019). Feature subset selection, hyperparameter tuning, and overall framework optimization were all performed with respect to maximizing weighted accuracy. The model development process is represented in Fig. 5.

Moreover, to assess the significance and robustness of the derived models and to ensure that neither selection bias nor overtraining influenced the results, 1,000 permutation tests were conducted. Each model's performance on the actual data was compared to its performance on randomly shuffled (permuted) labels to determine whether it significantly outperformed chance. Then, an empirical distribution of classification accuracies was generated, and a p-value was calculated as the probability that the accuracy from the permuted data exceeded that of the classifier on the actual data (Golland et al. 2005). Finally, to validate whether the selected features exhibit advanced discriminative ability in a global manner (i.e., the attainment of classification performance and robustness despite the classifier employed), the optimum feature set was chosen to achieve a statistically significant accuracy for each one of the tested ML models (Varma and Simon 2006; Ojala and Garriga 2009). Following the same workflow, the best-performing EV features were fused with EHR data in order to enhance the diagnostic capabilities of the framework. The feature selection and classification models were implemented using Python 3.9.18.

3 Results

3.1 Classification performance

Table 3 shows the classification performance of LR, SVM, RF, SGD, and DT models in differentiating stroke from mimics based on the 9 most important EV-related features from the initial 17 on the validation set, as defined by the relief feature selection method. The DT model outperformed the other models and achieved a weighted accuracy of 0.703 (accuracy = 0.81; precision = 0.85; recall = 0.91; f1-score = 0.88; ROC-AUC = 0.70). Regarding the statistical significance assessed through permutation testing, all models achieved p-values below 0.05 across 1,000 permutations, confirming the robustness of the results and highlights the consistent discriminative quality of the features used for stroke occurrence detection.

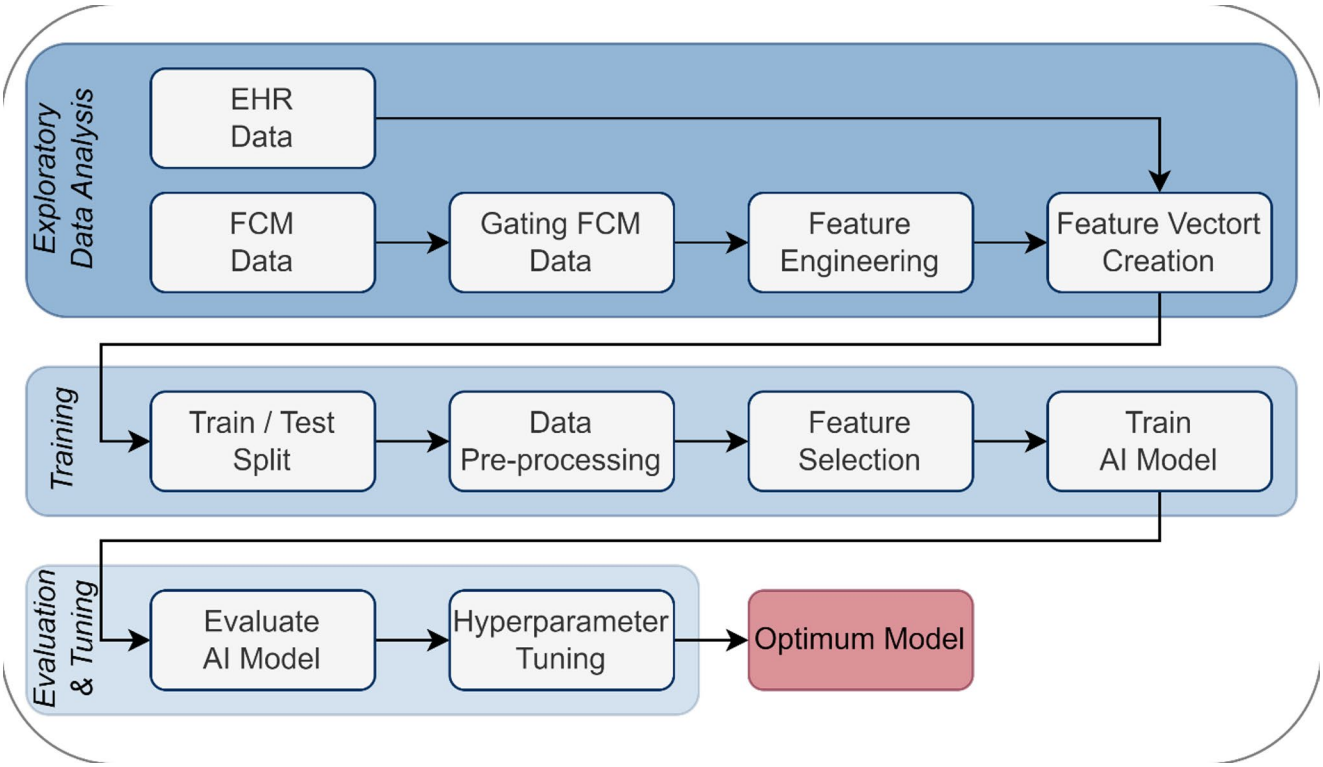

Fig. 5 Machine Learning model development workflow for stroke subtype classification using extracellular vesicles data acquired from flow-cytometry measurements and combination with EHR information. All preprocessing and feature selection occur inside inner folds of 5-fold cross-validation cross-validation, which reduces selection bias when tuning models and choosing features; label-permutation tests assess whether performance exceeds chance given the same data structure

Table 3 Models' performance on the Lact-FITC + flowcytometry measurements for stroke occurrence detection using only extracellular vesicles-related features

Classification Model	Precision	Recall	F1-score	ROC-AUC	Accuracy	Weighted Accuracy	*p*-value
Logistic Regression	0.58	0.62	0.55	**0.74**	0.57	0.62	*
SVM	0.57	0.60	0.55	0.69	0.60	0.60	*
Stochastic Gradient Descent	**0.91**	0.65	0.68	0.73	**0.83**	0.65	*
Random Forest	0.58	0.61	0.57	0.74	0.62	0.61	**
Decision Tree	0.85	**0.91**	**0.88**	0.70	0.81	**0.70**	***

* $p<0:05$; ** $p<0.01$; *** $p<0.001$, 1000 permutations

Table 4 Models' performance on the Lact-FITC + flowcytometry measurements for stroke occurrence detection using fusion of extracellular vesicles-related features and EHR information

Classification Model	Precision	Recall	F1-score	ROC-AUC	Unweighted Accuracy	Weighted Accuracy	*p*-value
Logistic Regression	0.69	0.76	0.67	0.82	0.69	0.76	*
SVM	0.66	0.72	0.60	0.78	0.62	0.72	**
Stochastic Gradient Descent	0.65	0.71	0.64	0.80	0.67	0.71	*
Random Forest	**0.93**	0.75	0.80	0.78	**0.88**	0.75	**
Decision Tree	0.78	**0.86**	**0.80**	**0.86**	0.83	**0.86**	***

* $p<0:05$; ** $p<0.01$; *** $p<0.001$, 1000 permutations

Following the EV-specific classification, features were fused with EHR data in order to enhance the models' performance in stroke occurrence detection. Thus, data regarding the sex, age, RR syst., RR diast., and thrombocytes were combined with the 9 most important EV features to form a total feature set to feed into the classification models. The performance of the fusion classification is resented in Table 4. Once again, the DT classifier achieved the highest weighted accuracy at 0.86 (accuracy = 0.83; precision = 0.78; recall = 0.86; F1-score = 0.88; ROC-AUC = 0.86). All models obtained p-values below 0.05 across 1,000 permutation

tests, further confirming the statistical robustness of the results. Fig. 6

3.2 Feature selection

The optimal fusion feature subset (obtained after feature selection), included both biomarker-based data (EV diameter distribution kurtosis, entropy, Std, skewness, mean, median, concentration of EVs with a diameter between 100 and 200 nm in the sample, total number of EVs and their concentration in the sample), and EHR information (sex, age, RR syst., RR diast., and thrombocytes). To assess whether differences between features across the two classes were statistically significant, a student's t-test was conducted, with the corresponding results shown in Fig. 7.

4 Discussion

In this study, we introduced an AI-driven DSS framework capable of distinguishing true stroke events from stroke mimics using a combination of EV biomarkers, biochemical and demographic data. The findings of this study hold significant implications for the future of stroke diagnosis, particularly in the context of emergency medical care. The implications of these findings are further discussed below.

4.1 Classification performance

Among the evaluated ML models, the DT classifier consistently outperformed the others across nearly all metrics, indicating strong potential for integration into clinical workflows as a pre-screening tool. Its balanced precision and recall (reflected in an F1-score of 0.80) highlights its suitability for stroke diagnosis, where both false positives and false negatives can adversely impact treatment decisions. Additionally, its ROC-AUC and accuracy values of 0.86 demonstrate strong discriminative ability between stroke and stroke-mimic cases, while the statistically significant p-value (<0.001) in permutation testing support its robustness and reliability.

4.2 Selected features

As mentioned earlier, the optimal feature subset for stroke detection comprised both EV-derived and EHR-based variables. Notably, the subset included predominately EV features (9 of the 14) originate from EV measurements. Although most of these features did not exhibit strong

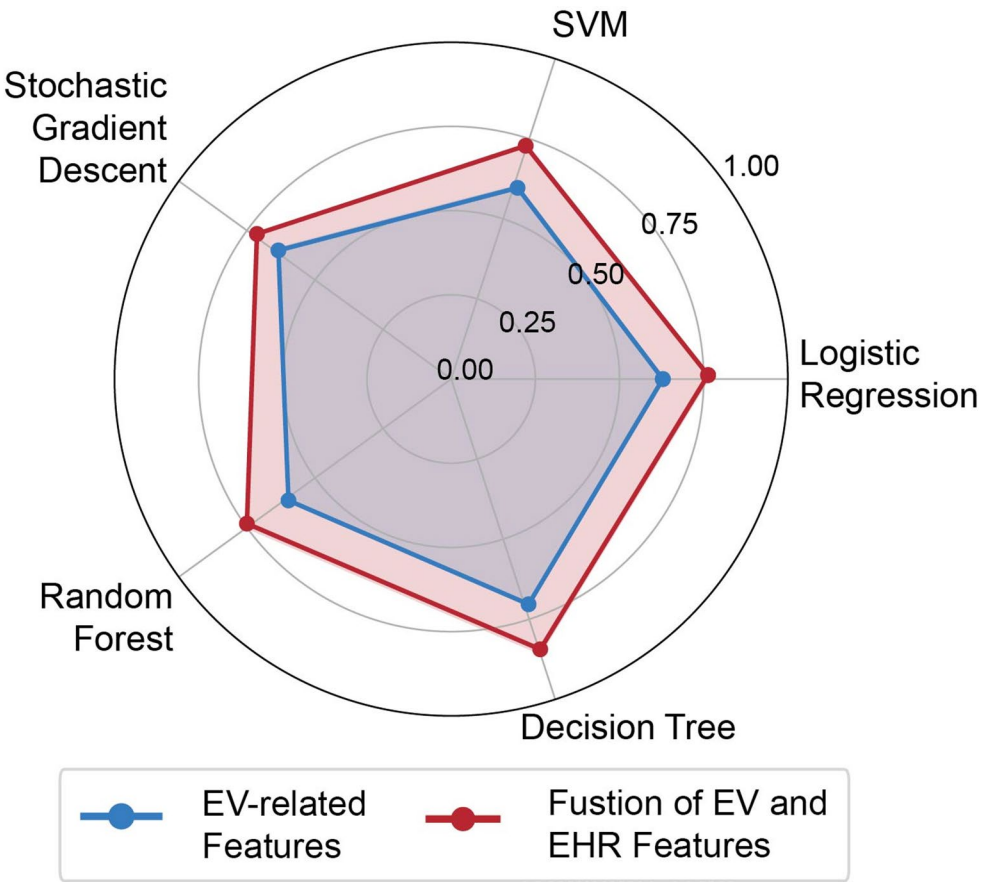

Fig. 6 Machine Learning models' weighted accuracy performance of stroke occurrence classification, using only extracellular vesicles-related features, and combining them with EHR data. The data fusion outperforms single-modality baselines, consistent with literature that single blood panels alone often yield moderate accuracy

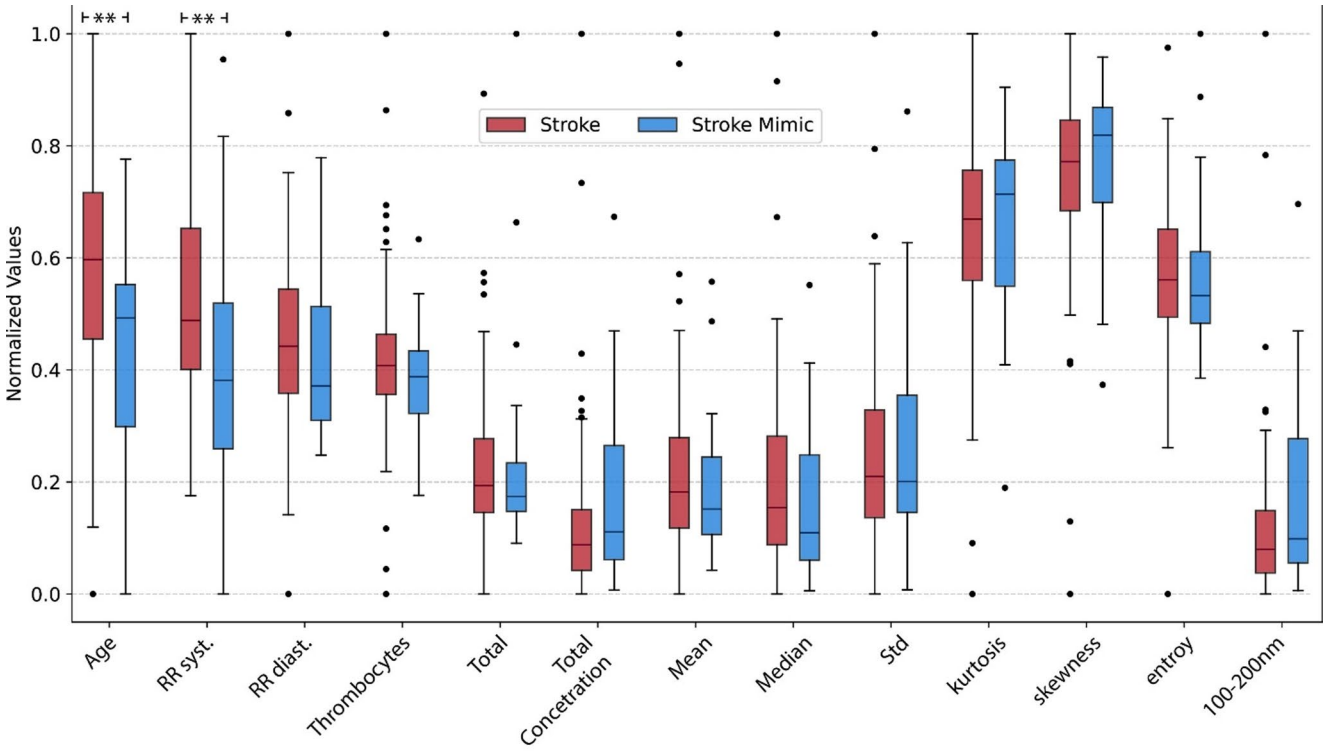

Fig. 7 Student's t-test analysis of global features. Boxplots indicate the median, the first and third quartiles. Whiskers indicate the smallest and the largest values. Open circles indicate outliers. The limited statistical significance derived from Student's t-test analysis supports the Machine Learning integration. ** indicates student's t-test statistical significance of $p < 0.01$

statistical significance, they still provide valuable insights into their potential relevance in the diagnostic process. As such, the selected EV features offer a novel perspective for stroke diagnosis. Metrics such as the mean and median EV diameter, total EV count, and overall concentration, are aligned with previous studies, serving as key indicators of their biological and biophysical properties (Stenz et al. 2020b). Particularly notable is the concentration of EVs within the 100–200 nm range, as these vesicles are predominantly derived from platelets and endothelial cells (cell types directly involved in coagulation and inflammatory pathways that play central roles in stroke pathophysiology) (Taus et al. 2019). Furthermore, statistical descriptors of the EV diameter distribution (including standard deviation, entropy, kurtosis, and skewness) capture the variability and extremities in EV size profiles within plasma, and thus indicating complex, non-linear relationships between such variables that may not be detectable using conventional univariate statistical analysis.

Regarding EHR biochemical features, several variables (despite having pure statistical differences) are strongly linked to stroke pathogenesis as indicated by previous research. For instance, thrombocyte count is implicated in the inflammatory and thrombotic processes underlying stroke development (Elkind 2009) (Bi et al. 2021). Similarly, systolic and diastolic blood pressure are well-established risk factors for cerebrovascular disease, diseases as they contribute to endothelial damage, increasing the risk of atherothrombotic events (Lattanzi et al. 2021). Notably, the systolic blood pressure feature showed a statistically significant difference between the stroke and stroke-mimic groups, further supporting its relevance in the diagnostic context. Concerning the EHR demographic features, the student's t-test indicated a highly significant difference in age between stroke and non-stroke groups, making age as one of the strongest predictors of stroke occurrence. In fact, age is a well-established risk factor for stroke, with older individuals exhibiting increased susceptibility to both ischemic and hemorrhagic strokes due to age-related vascular changes (Kelly-Hayes 2010). Sex was also selected as an important characteristic for stroke classification. To assess its influence as a global predictor, additional analysis was employed by removing sex from the optimal feature subset, retraining and re-evaluating the models. The resulting performance showed poorer results with a nearly 20% reduction in classification accuracy, demonstrating that excluding sex substantially weakens predictive power. This highlights the importance of including sex in the modeling processes, as it reflects real-world clinical conditions in which sex plays a critical role in shaping disease risk and progression.

4.3 Future considerations

The results obtained in this study demonstrate that the selected features (derived from both EV measurements and patient biochemical and demographic data) yield high classification performance in diagnosing stroke occurrence. However, caution is required when interpreting the biological significance of these features. The primary concern is that the feature selection process focuses on improving classification performance rather than establishing direct associations between the selected features and the underlying pathophysiological mechanisms of stroke. Therefore, some features might have been included primarily for their ability to reduce noise or enhance performance rather than due to their inherent relevance to stroke pathology. Moreover, although the classification models exhibit robust performance, variability between patients may influence the overall results. Factors such as individual differences in biochemical profiles, EV release patterns, and responses to stroke could contribute to this variability. Further research should be conducted to validate the model's generalizability across diverse populations and to explore the physiological implications of the selected features in more depth. In addition to these considerations, another major limitation of the current approach is the time-consuming nature of blood sample preparation and EV analysis, which may restrict its real-time applicability in emergency settings. Addressing this challenge will require future work focused on streamlining the analytical workflow or developing alternative, faster diagnostic methods for photonics-based EV assessment without compromising accuracy. Furthermore, the scope of the framework could be expanded into a multiclass triage aid system, which not only identifies the stroke occurrence among stroke mimic conditions, but also classifies the specific stroke type, such as ischemic, hemorrhagic strokes, TIA, directing the patient to the appropriate clinic. Regarding the AI models utilized, deep learning architectures could be incorporated in addition to the current methods, once larger patient cohorts become available. Despite these limitations, the present framework shows significant potential in reducing diagnosis time after the blood acquisition and processing and improving the accuracy of prompt stroke detection. The system's ability to discern stroke cases from other conditions with similar symptoms represents a crucial advancement in clinical decision-making, potentially expediting the appropriate treatment of stroke patients.

5 Conclusion

In this paper, we investigated the integration of novel blood-based biomarkers (i.e., extracellular vesicles, EVs extracted via flow cytometry) with patient demographic and biochemical information to enhance the early and accurate diagnosis of stroke. Our approach focuses on distinguishing true stroke cases from other conditions presenting with similar symptoms, thereby reducing diagnostic delays and supporting timely patient transport to appropriate treatment facilities. On this premise, the proposed framework achieved a classification accuracy of 0.86, using a small feature subset whose robustness and clinical reliability were confirmed through statistical testing. Although certain features did not exhibit individual statistical differences between the two groups, their combined contribution within the model captures the multifactorial nature of stroke, ultimately improving diagnostic performance. Inherently interpretable ML algorithms are incorporated, as they are considered suitable for medium sized clinical cohorts and for high-stakes decision-making. To limit overfitting and validate the robustness of the results, k-fold cross validation was performed during the training and hyperparameter tuning process, as well as label-permutation testing. Future steps include the expansion of the research to external datasets for the further validation of the framework's performance and the most important features analysis. The results indicate that single-particle EV size distribution features are able to encode discriminative signal, and fusion with minimal EHR variables improves performance. In addition, interpretable models, such as the DT algorithm are sufficient and preferable for high-stakes clinical use. Overall, the findings demonstrate the potential of AI-driven DSS to assist in the prompt diagnosis of stroke, providing an essential tool that complements existing diagnostic methods, significantly improving patient care. This work highlights the value of integrating advanced biomarkers, clinical data, and AI methodologies, combining standard diagnostics with novel photonics methodologies, to develop a more comprehensive diagnostic framework capable of addressing the challenges of time-sensitive conditions such as stroke.

Acknowledgements We would like to express our sincere gratitude to Yvo B. W. E. M. Roos, Dr. Nyika Kruyt, and A. Yael Nossent for their valuable support in establishing the Circulating Nano Traces to Identify the Cause of Stroke II (CINTICS II) study, through which the data used in this work were collected and made available.

Author contributions All authors contributed to the writing of the original draft, review & editing of this study. Additional contributions included: Conceptualization: Vasileios E. Katsigiannis and George K. Matsopoulos; Data curation: Naomi C. Buntsma, Joyce Rops and Edwin van der Pol; Formal analysis: Ourania Petropoulou, Ioannis A. Vezakis and Dimitrios I. Fotiadis; Funding acquisition: George K. Matsopoulos; Investigation: Edwin van der Pol, Ioannis Kakkos,

Ioannis A. Vezakis and Dimitrios I. Fotiadis; Methodology: Stavros-Theofanis Miloulis, Vasileios E. Katsigiannis, and Ioannis Kakkos; Project administration: George K. Matsopoulos; Resources: Naomi C. Buntsma and Joyce Rops; Software: Vasileios E. Katsigiannis, Stavros-Theofanis Miloulis and Ioannis A. Vezakis; Supervision: George K. Matsopoulos; Validation: Ourania Petropoulou and Dimitrios I. Fotiadis; Visualization: Vasileios E. Katsigiannis and Ioannis Kakkos.

Funding This research is supported by the PHOREVER project funded by the European Union under the Horizon EU research and innovation program (GA ID: 101093171). This research was also partially supported by the Dutch Heart Foundation, research grant CINTICS 2018B031, Impulse Grant CINTICS II (01-002-2023-0142) and the Dutch Research Council (NWO) for VIDI project 19724.

Declarations

Competing interests Edwin van der Pol is co-founder and shareholder of Exometry BV. The other authors declare no competing interests.

Ethical approval The CINTICS II study was approved by the Ethical Review Board of Amsterdam University Medical Centers, location AMC, with a deferred consent process (approval number NL72929.018.20).

References

Bates DW, Saria S, Ohno-Machado L et al (2014) Big Data In Health Care: using analytics to identify and manage high-risk and high-cost patients. Health Aff 33:1123–1131. https://doi.org/10.1377/hlthaff.2014.0041

Benjamin EJ, Virani SS, Callaway CW et al (2018) Heart disease and stroke statistics—2018 update: a report from the American Heart Association. Circulation. https://doi.org/10.1161/CIR.0000000000000558

Bi R, Chen S, Chen S et al (2021) The role of leukocytes in acute ischemic stroke-related thrombosis: a notable but neglected topic. Cell Mol Life Sci 78:6251–6264. https://doi.org/10.1007/s00018-021-03897-5

Bisong E (2019) Logistic regression. Building machine learning and deep learning models on Google cloud platform. A, Berkeley, CA, pp 243–250

Buntsma NC, Shahsavari M, Gąsecka A et al (2023) Preventing swarm detection in extracellular vesicle flow cytometry: a clinically applicable procedure. Res Pract Thromb Haemost 7:100171. https://doi.org/10.1016/j.rpth.2023.100171

Buntsma NC, Hau CM, Los M et al (2025) Identification of ischemic stroke patients based on plasma concentrations of extracellular vesicles. Transl Stroke Res 16:2082–2092. https://doi.org/10.1007/s12975-025-01371-z

Candelaresi P, Di Monaco C, Pisano E (2023) Stroke chameleons: diagnostic challenges. Eur J Radiol Open 11:100533. https://doi.org/10.1016/j.ejro.2023.100533

Casado-Fernández L, Laso-García F, Piniella D et al (2024) The proteomic signature of circulating extracellular vesicles following intracerebral hemorrhage: novel insights into mechanisms underlying recovery. Neurobiol Dis 201:106665. https://doi.org/10.1016/j.nbd.2024.106665

Cervantes J, Garcia-Lamont F, Rodríguez-Mazahua L, Lopez A (2020) A comprehensive survey on support vector machine classification: applications, challenges and trends. Neurocomputing 408:189–215. https://doi.org/10.1016/j.neucom.2019.10.118

Charbuty B, Abdulazeez A (2021) Classification based on decision tree algorithm for machine learning. JASTT 2:20–28. https://doi.org/10.38094/jastt20165

Couch Y, Akbar N, Davis S et al (2017) Inflammatory stroke extracellular vesicles induce macrophage activation. Stroke 48:2292–2296. https://doi.org/10.1161/STROKEAHA.117.017236

Czap AL, Bahr-Hosseini M, Singh N et al (2022) Machine learning automated detection of large vessel occlusion from mobile stroke unit computed tomography angiography. Stroke 53:1651–1656. https://doi.org/10.1161/STROKEAHA.121.036091

Elkind MSV (2009) Inflammatory markers and stroke. Curr Cardiol Rep 11:12–20. https://doi.org/10.1007/s11886-009-0003-2

Gkantzios A, Kokkotis C, Tsiptsios D et al (2023) Evaluation of blood biomarkers and parameters for the prediction of stroke survivors' functional outcome upon discharge utilizing explainable machine learning. Diagnostics 13:532. https://doi.org/10.3390/diagnostics13030532

Golland P, Liang F, Mukherjee S, Panchenko D (2005) Permutation tests for classification. In: Auer P, Meir R (eds) Learning theory. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 501–515

Heo J, Yoon JG, Park H et al (2019) Machine learning–based model for prediction of outcomes in acute stroke. Stroke 50:1263–1265. https://doi.org/10.1161/STROKEAHA.118.024293

Jadhav AP, Desai SM, Liebeskind DS, Wechsler LR (2020) Neuroimaging of acute stroke. Neurol Clin 38:185–199. https://doi.org/10.1016/j.ncl.2019.09.004

Kamtchum-Tatuene J, Jickling GC (2019) Blood biomarkers for stroke diagnosis and management. Neuromol Med 21:344–368. https://doi.org/10.1007/s12017-019-08530-0

Kelly-Hayes M (2010) Influence of age and health behaviors on stroke risk: lessons from longitudinal studies. J Am Geriatr Soc. https://doi.org/10.1111/j.1532-5415.2010.02915.x. 58:

Kim SJ, Moon GJ, Bang OY (2013) Biomarkers for stroke. J Stroke 15:27. https://doi.org/10.5853/jos.2013.15.1.27

Kim TY, Park JY, Cho YS et al (2025) Identification of stroke biomarkers using proteomic profiling of extracellular vesicles derived from human blood: a preliminary study. Front Neurol. https://doi.org/10.3389/fneur.2025.1587389

Lacroix R, Judicone C, Poncelet P et al (2012) Impact of pre-analytical parameters on the measurement of circulating microparticles: towards standardization of protocol. J Thromb Haemost 10:437–446. https://doi.org/10.1111/j.1538-7836.2011.04610.x

Lacroix R, Judicone C, Mooberry M et al (2013) Standardization of pre-analytical variables in plasma microparticle determination: results of the International Society on Thrombosis and Haemostasis SSC collaborative workshop. J Thromb Haemost 11:1190–1193. https://doi.org/10.1111/jth.12207

Lattanzi S, Brigo F, Silvestrini M (2021) Blood pressure variability and stroke: a risk marker of outcome and target for intervention. J Clin Hypertens 23:103–105. https://doi.org/10.1111/jch.14092

Leppälä JM, Virtamo J, Fogelholm R et al (1999) Different risk factors for different stroke subtypes: association of blood pressure,

cholesterol, and antioxidants. Stroke 30:2535–2540. https://doi.org/10.1161/01.str.30.12.2535
Li F, Li L, Peng R et al (2024) Brain-derived extracellular vesicles mediate systemic coagulopathy and inflammation after traumatic brain injury. Int Immunopharmacol 130:111674. https://doi.org/10.1016/j.intimp.2024.111674
Liu S, Jiang H, Liu L, Han J (2018) Gradient descent using stochastic circuits for efficient training of learning machines. IEEE Trans Comput Aided Des Integr Circuits Syst 37:2530–2541. https://doi.org/10.1109/tcad.2018.2858363
Liu J, Chou EL, Lau KK et al (2022) Machine learning algorithms identify demographics, dietary features, and blood biomarkers associated with stroke records. J Neurol Sci 440:120335. https://doi.org/10.1016/j.jns.2022.120335
Mainali S, Darsie ME, Smetana KS (2021) Machine learning in action: stroke diagnosis and outcome prediction. Front Neurol 12:734345. https://doi.org/10.3389/fneur.2021.734345
Martinez-Gutierrez JC, Kim Y, Salazar-Marioni S et al (2023) Automated large vessel occlusion detection software and thrombectomy treatment times. JAMA Neurol 80:1182–1190. https://doi.org/10.1001/jamaneurol.2023.3206
Meyers PM, Schumacher HC, Connolly ES et al (2011) Current status of endovascular stroke treatment. Circulation 123:2591–2601. https://doi.org/10.1161/CIRCULATIONAHA.110.971564
Ojala M, Garriga GC (2009) Permutation Tests for Studying Classifier Performance. In: 2009 Ninth IEEE International Conference on Data Mining. IEEE, Miami Beach, FL, USA, pp 908–913
Patil S, Rossi R, Jabrah D, Doyle K (2022) Detection, diagnosis and treatment of acute ischemic stroke: current and future perspectives. Front Med Technol 4:748949. https://doi.org/10.3389/fmedt.2022.748949
Rahmig J, Chanpura A, Schultz A et al (2024) Blood-based protein biomarkers during the acute ischemic stroke treatment window: a systematic review. Front Neurol 15:1411307. https://doi.org/10.3389/fneur.2024.1411307
Reymond S, Gruaz L, Schvartz D et al (2025) Proteomic analysis of plasma-derived extracellular vesicles: insights into acute stroke pathophysiology. J Proteomics 319:105468. https://doi.org/10.1016/j.jprot.2025.105468
Robnik-Šikonja M, Kononenko I (2003) Theoretical and empirical analysis of relieff and RReliefF. Mach Learn 53:23–69. https://doi.org/10.1023/A:1025667309714
Rudin C (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 1:206–215. https://doi.org/10.1038/s42256-019-0048-x
Shah KH, Edlow JA (2004) Transient ischemic attack: review for the emergency physician. Ann Emerg Med 43:592–604. https://doi.org/10.1016/j.annemergmed.2003.12.025
Shaik AB, Srinivasan S (2019) A Brief Survey on Random Forest Ensembles in Classification Model. In: Bhattacharyya S, Hassanien AE, Gupta D, (eds) International Conference on Innovative Computing and Communications. Springer Singapore, Singapore, pp 253–260
Stenz KT, Just J, Blauenfeldt RA, Drasbek KR (2020) Extracellular vesicles in acute stroke diagnostics. Biomedicines 8:248. https://doi.org/10.3390/biomedicines8080248
Taus F, Meneguzzi A, Castelli M, Minuz P (2019) Platelet-derived extracellular vesicles as target of antiplatelet agents. What is the evidence? Front Pharmacol 10:1256. https://doi.org/10.3389/fphar.2019.01256
Théry C, Witwer KW, Aikawa E et al (2018) Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. J of Extracellular Vesicle 7:1535750. https://doi.org/10.1080/20013078.2018.1535750
Thulin Å, Lindbäck J, Granger CB et al (2020) Extracellular vesicles in atrial fibrillation and stroke. Thromb Res 193:180–189. https://doi.org/10.1016/j.thromres.2020.07.029
Topol EJ (2019) High-performance medicine: the convergence of human and artificial intelligence. Nat Med 25:44–56. https://doi.org/10.1038/s41591-018-0300-7
Urbanowicz RJ, Meeker M, La Cava W et al (2018) Relief-based feature selection: introduction and review. J Biomed Inform 85:189–203. https://doi.org/10.1016/j.jbi.2018.07.014
Varma S, Simon R (2006) Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics 7:91. https://doi.org/10.1186/1471-2105-7-91
Vujovic ŽĐ (2021) Classification model evaluation metrics. Int J Adv Comput Sci Appl (IJACSA) 12. https://doi.org/10.14569/IJACSA.2021.0120670
Welsh JA, Arkesteijn GJA, Bremer M et al (2023) A compendium of single extracellular vesicle flow cytometry. J Extracell Vesicles 12:e12299. https://doi.org/10.1002/jev2.12299
Widimsky P, Snyder K, Sulzenko J et al (2023) Acute ischaemic stroke: recent advances in reperfusion treatment. Eur Heart J 44:1205–1215. https://doi.org/10.1093/eurheartj/ehac684
Wilson D, Ambler G, Shakeshaft C et al (2018) Cerebral microbleeds and intracranial haemorrhage risk in patients anticoagulated for atrial fibrillation after acute ischaemic stroke or transient ischaemic attack (CROMIS-2): a multicentre observational cohort study. Lancet Neurol 17:539–547. https://doi.org/10.1016/s1474-4422(18)30145-5
Yan LL, Li C, Chen J et al (2016) Prevention, management, and rehabilitation of stroke in low- and middle-income countries. eNeurolSci 2:21–30. https://doi.org/10.1016/j.ensci.2016.02.011
Yang Y, Boza-Serrano A, Dunning CJR et al (2018) Inflammation leads to distinct populations of extracellular vesicles from microglia. J Neuroinflammation 15:168. https://doi.org/10.1186/s12974-018-1204-7
Yuana Y, Sturk A, Nieuwland R (2013) Extracellular vesicles in physiological and pathological conditions. Blood Rev 27:31–39. https://doi.org/10.1016/j.blre.2012.12.002