LUNG CANCER DETECTION THROUGH HYBRID ALGORITHMS

S. Sulaiha Beevi,
Research Scholar, Bharathiar University, Coimbatore. E-mail: sulaiha2010@gmail.com
Dr. K.L. Shunmuganathan,
Director, Jaya Engineering College, Chennai. Email: klsnathan@gmail.com
Dr. Baskaran Sriramulu,
Assistant Professor, Birla Institute of Technology & Science, Pilani, Dubai Campus, Email: baskaran@dubai.bits-pilani.ac.in

ABSTRACT

Nowadays, lifestyle disorders such as hypertension, diabetes, cancer, and heart disease lead to many diseases in humans, making accurate disease prediction through symptoms difficult for physicians. Early-stage disease prediction is a challenging task, but data mining can play an important role in overcoming this problem. Machine learning algorithms can uncover hidden information from disease datasets. In this paper, the study focuses on predicting lung cancer at an early stage through a multi-agent based architecture using machine learning. Three different methods are proposed in this paper for predicting lung cancer at an early stage. Initially, lung cancer is predicted through Logistic Regression (LR), k Nearest Neighbour (KNN), Naive Bayes (NB), and Support Vector Machine (SVM) methods. The second method applies machine learning in a multi-agent platform for predicting lung cancer at an early stage using a Java Agent Development Environment based Multi Agent System (JADE-MAS). The JADE-MAS consists of three agents, the Medical practitioner Agent, Classifier Agent, and Database Agent. The third method applies JADE-MAS with novel hybrid classifiers such as Multiple Linear Regression + k Nearest Neighbour (MLR+KNN), Gaussian Kernel Support Vector Machine + Linear Regression (GKSVM+LR), and Gaussian Kernel Support Vector Machine + k Nearest Neighbour (GKSVM+KNN) for lung cancer prediction at an early stage. The analysis focuses on agent-based and non-agent based classification for lung cancer prediction at an early stage using sensitivity, specificity, and accuracy. The optimum method for detecting lung cancer at an early stage is identified from the analysis, and the advantages and disadvantages of each method are studied. It is suggested that the use of feature selection algorithms and optimization approaches can further improve the efficacy of these predictive classifiers in diagnosing diseases, and developing a multi-agent system for the DM system using SPADE as a platform is another area of interest.

Keywords: Lung cancer; machine learning; JADE-MAS

INTRODUCTION

Data Mining (DM) is utilized in various fields, such as marketing, engineering, medical, web mining, mobile computing, and expert prediction. In healthcare, DM is employed to identify fraud and abuse, and to improve clinical decisions by using data modelling and analysis. Failure to utilize collected data can result in improper treatment, diagnosis, adverse drug effects, and high medical costs. This chapter examines how DM can be used to diagnose lung cancer and the need for an agent system to predict lung cancer. Cancer occurs when gene changes lead to uncontrolled cell division, resulting in the formation of tumours or abnormal blood cells. Cancer cells divide uncontrollably and never stop creating new cells, which can spread to other parts of the body. Cancer is caused by mistakes in cell division or DNA damage, such as exposure to radiation and compounds found in cigarette smoke. Each person's cancer has a unique genetic makeup, with different genetic alterations in tumour cells. Abnormal growth or mutations of cancer cells are called "driver mutations," while normal mutations are "passenger mutations." Tumour genes can prevent cell growth by slowing down mutations, resulting in damaged DNA or cell death. Lung cancer arises from cells lining the bronchi, bronchioles, or alveoli of the lungs, which are responsible for gas exchange by taking in oxygen and releasing carbon dioxide. The trachea carries air through the mouth or nose and divides into bronchi and bronchioles that lead to the alveoli. The lungs filter inhaled air by bronchi tubes, and red blood cells absorb carbon dioxide and transport oxygen from the lungs to the blood. Breathing is controlled by the diaphragm, a dome-shaped muscle below the lungs that flattens during inhalation and extends during exhalation. Adults typically breathe between 12 and 20 times per minute, increasing to 45 per minute during vigorous activity. Lung cancer is more common in smokers and occurs in the right lung, which is larger than the left and located below the heart. There are two main types of lung cancer: NSCLC and SCLC. SCLC is more common in chain smokers, while NSCLC includes squamous cell carcinoma, adenocarcinoma, and large cell carcinoma. Agent technology is a rapidly growing area in AI. A software agent is a piece of software that operates independently in a specific environment on behalf of a user or another program. Agents are autonomous entities that display characteristics of learning, cooperation, and mobility, and can act proactively and responsively. The environment and agents together form an AI system where agents behave and may interact with other agents or change the environment using sensors and actuators. In FIPA, agents are closed software entities that can interact and communicate with humans, other agents, and legacy systems.

Multi-agent systems (MAS) are clusters of agents that work on decentralized, unknown mediums to solve problems and install numerical inter-linkage mediums. MAS creates goal-oriented autonomous agents that cooperate and communicate commonly, which is useful in distributed data mining (DDM). DDMs rely on goal-oriented processes, distributed clustering, and predictions to allocate network resources. MAS can handle complex applications that demand distributed solutions to problems. MAS-based DDMs allow agents to build their own personal learning models and control acquirable knowledge or results. However, these systems have drawbacks like the inability of local agents to exchange information, the difficulty in developing capabilities of autonomous agents, and the finding of private and non-shareable material.Agents with their environments are components of MAS. Agents in MAS can be categorized as passive agents, active agents, and cognitive agents based on their objectives. Agent settings can be shaped by characteristics like accessibility, determinism, dynamics, discreteness, episodicity, and dimensionality. Middleware manages agent actions in MAS and offers high levels of design abstractions that control resource accesses and agent co-ordinations. MAS has been successfully used in robotics, oceanography, medicines, e-commerce, and other industries. MAS can address many software engineering challenges, including extensions, reusability, and robustness.FIPA is an international organization that develops standards for generic agent technologies to create sophisticated systems with excellent interoperability. The FIPA standard defines a reference model for an agent platform and a list of required services. The FIPA Agent Platform consists of several components, including the AMS, DF, ACC, and MTS. The AMS controls access and use of the platform, while the DF offers yellow pages services for agents to find out about other agents' services. The ACC supports both agent administration and inter-agent communication, and the MTS securely sends FIPA ACL messages between agents.

PROBLEM STATEMENT

Lung data set-based data mining helps physicians discover hidden information about cancer. However, traditional algorithms such as SVM, KNN, LR, and NB have limitations. Researchers should focus on agent-based data collection for lung diseases, as symptoms vary by region and lifestyle. Hybrid algorithms are proposed to solve these limitations.Multi-agent system (MAS) is used in machine learning for decision-making and achieving common goals. However, there is a need for a new classifier as some algorithms such as SVM and KNN. SVM is not suitable for large datasets due to its training process and poor performance on imbalanced datasets. KNN suffers from high computational costs and requires feature scaling.Logistic regression predicts outcomes based on independent features and may not accurately forecast test results. Naive Bayes assumes all features are independent, limiting its real-world application and suffers from the 'zero-frequency problem.'To address these issues, hybrid algorithms with multi-agent data collection have been proposed.

Objectives

This research aims to develop an intelligent data mining system for lung databases using a multi-agent system. The best performing hybrid combinations were selected and are presented in this thesis. The specific objectives are as follows:

♦ To investigate lung cancer disease patterns using both standard data sets and agent-based data collection.
♦ To use a multi-agent system for early detection of lung cancers in the data set.
♦ To propose hybrid algorithms for the lung data set, specifically LR + KNN, GKSVM + LR, and GKSVM + KNN.
♦ To evaluate the performance of the proposed hybrid algorithms using precision, recall, and true negative measures on both standard data sets and agent-based lung cancer data sets.
♦ To analyse the advantages and disadvantages of the proposed method for both standard data sets and agent-based lung cancer data sets.

LITERATURE SURVEY

Reference	Methods	Disadvantages
[1]	NB	The Naive Bayes method's primary disadvantage is its assumption of
[2]	SVM	SVM may not perform well when there is a significant amount of noise or overlapping target classes in the dataset. Additionally, if the number of features for each data point exceeds the number of training data samples, SVM's performance may not meet expectations.
[3]	ANN	Artificial neural networks require processors capable of parallel pro-
[4]	XGBoost	XGBoost's performance may suffer when applied to sparse and unstructured data. Moreover, before feeding data into the models, manual creation of dummy variables and label encodings for categorical features is required in XGBoost.
[5]	Rough set theory	A significant computational time requirement is a major drawback of

Reference No	Methods	Disadvantages
[7]	ANFIS	ANFIS has several drawbacks, including high computational costs, reduced interpretability with larger inputs, susceptibility to the curse of dimensionality, and difficulty in selecting appropriate membership functions.
[8]	NLP	Natural language processing (NLP) systems are limited to a single application and are unable to adapt to new problem areas or domains. Additionally, the user interface of an NLP system lacks features that could enhance user interaction.
[9]	Extreme Learning Machine	When approximating highly nonlinear data, extreme learning machines do not yield improved accuracy.
[10]	Fuzzy Set	Extensive testing with various equipment is required for approval and verification of a fuzzy information-based framework. Additionally, the imprecise reasoning in fuzzy sets may be mistaken for the probability hypothesis.
[11]	PSO	In high-dimensional space, getting stuck in a local optimum is common, and the iterative process of PSO has a slow convergence rate for finding solutions. Moreover, PSO's optimal solution has low local search ability.
[12]	Q Learning	Q Learning can only be applied in settings with discrete and finite state and action spaces. The use of a policy different from the target policy to explore actions in Q Learning results in unreliable policies
[13]	Decision tree	The decision tree algorithm can easily over fit as it lacks a mechanism to prevent it, resulting in complex decision rules. Furthermore, minor data alterations can cause significant changes in the tree structure, leading to inconsistency.
[14]	J48	J48 produces numerous nodes with zero values under certain conditions, which do not contribute to class formation for the classification task. These nodes increase the tree's complexity and reduce the decision tree's usability. The insignificant branches produced by this algorithm also contribute to overfitting.
[15]	Decision Stump	The decision stump model is unable to solve complex learning problems due to its single level of decision trees.

Inferences from literature survey

Medical professionals are being aided in the diagnosis of several illnesses due to the significant increase in efficiency of categorization and identification systems. This chapter investigated various classification approaches, including hybrid and agent-based methods, to identify numerous illnesses. The proposed system aims to predict multiple diseases using the three categories mentioned above. Based on the results, the agent-based system with a classifier outperforms the other methods.

METHODOLOGY

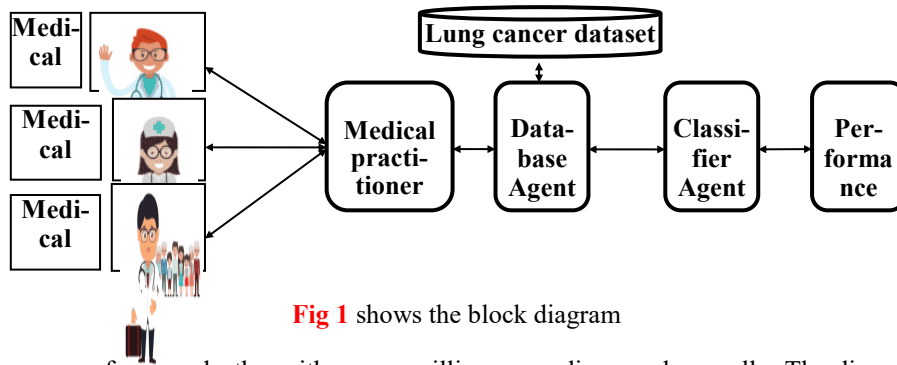**Figure 1** shows the block diagram of proposed algorithms.

Fig 1 shows the block diagram

Data Mining

Lung cancer is a leading cause of global deaths, with over a million cases diagnosed annually. The disease is characterized by the unrestrained proliferation of abnormal cells in the lungs, and there are three primary types: carcinoid, SCLC, and NSCLC. Early detection is crucial for effective treatment. DM is particularly important in healthcare, where it helps extract knowledge from large databases. The mining process involves classification, clustering, association rule mining, and prediction. A recent study examined the use of classification algorithms, such as LR, KNN, NB, and SVM, to predict lung cancer using an agentless system.ML is a branch of artificial intelligence that involves computer algorithms improving their performance over time by building a model out of training data. This approach is used in applications where classical algorithms are impractical, such as email filtering and computer vision. While ML and DM share many techniques, they differ in their goals. DM is focused on identifying undiscovered properties in data, while ML is focused on prediction based on known features learned from training data. ML has two main goals: categorizing data using established models and forecasting future results using these models.

Machine Learning

DM and ML classifiers are becoming increasingly important tools in medical diagnosis, reducing potential errors and assisting both expert and unskilled doctors. This study evaluates the performance of multi-agent-based ML hybrid algorithms for lung cancer diagnosis using classifier algorithms such as LR, KNN, NB, and SVM. The SVM classifier was found to be the most accurate and computationally efficient, producing better results in less time and with fewer resources than the other classifiers. The study also includes methods for altering agent behaviour, such as unlocking an agent through message reception, timeout, or restart. Once the agent process is completed, traditional classifiers are used for disease diagnosis.

Hybrid algorithms

The use of a single classifier for accurately diagnosing all illnesses has proven challenging despite the development and testing of numerous individual ML-based classifiers in recent years. To address this issue, researchers have proposed hybrid classification algorithms that optimize individual classifiers. In this study, three hybrid classifiers (GKSVM +LR, MLR+KNN, and GKSVM+KNN) were developed and evaluated for diagnosing lung cancer disease. The study compared the performance of these hybrid systems with current classifiers using parameters such as precision, recall, F-measure, and accuracy. The proposed hybrid system aims to deliver generalized performance over a wide range of benchmark data sets for multiple diseases. The study also discusses the use of hybrid approaches in various fields, including medicine, and highlights the challenges of using individual classifiers for disease diagnosis. The MLR+KNN technique is suggested for classifying lung cancer by effectively integrating the MLR and KNN algorithms to reduce classification errors. This hybrid classifier utilizes multiple regressions to extract label-dependent variables from the label space and considers label dependence in both the label and prediction spaces. The goal of this classifier is to divide data into several classes. The cosine similarity function is used to compute similarity between samples in the lung cancer data set. MLR is employed to extract data from the label space, and a linear model is created for each label using the training set's labels. Hybrid KNN is utilized to identify the k-nearest neighbors that are the most comparable. The two primary algorithms used in this multi-label classifier are MLR and KNN. The objective is to construct a hybrid classifier D that consists of a GKSVM classifier S and an LR classifier L for a set of training data X with m attributes. The GKSVM classifier utilizes the kernel technique to transform a low-dimensional input space into a higher-dimensional one to make it separable. A wide margin is viewed positively while a small margin is viewed negatively.

RESULTS AND DISCUSSIONS

The experimental results and accuracy score were evaluated using lung cancer data obtained from the UCI repository. The qualities were selected based on recommendations from medical practitioners at hospitals. MATLAB was utilized to analyse the lung cancer dataset, and the predicted result's effectiveness was assessed. Confusion matrices for four classifiers, namely NB, LR, KNN, and SVM, are presented in Table 1. Additionally, Table 2 displays the evaluation metrics versus JADE-MAS + classifiers.
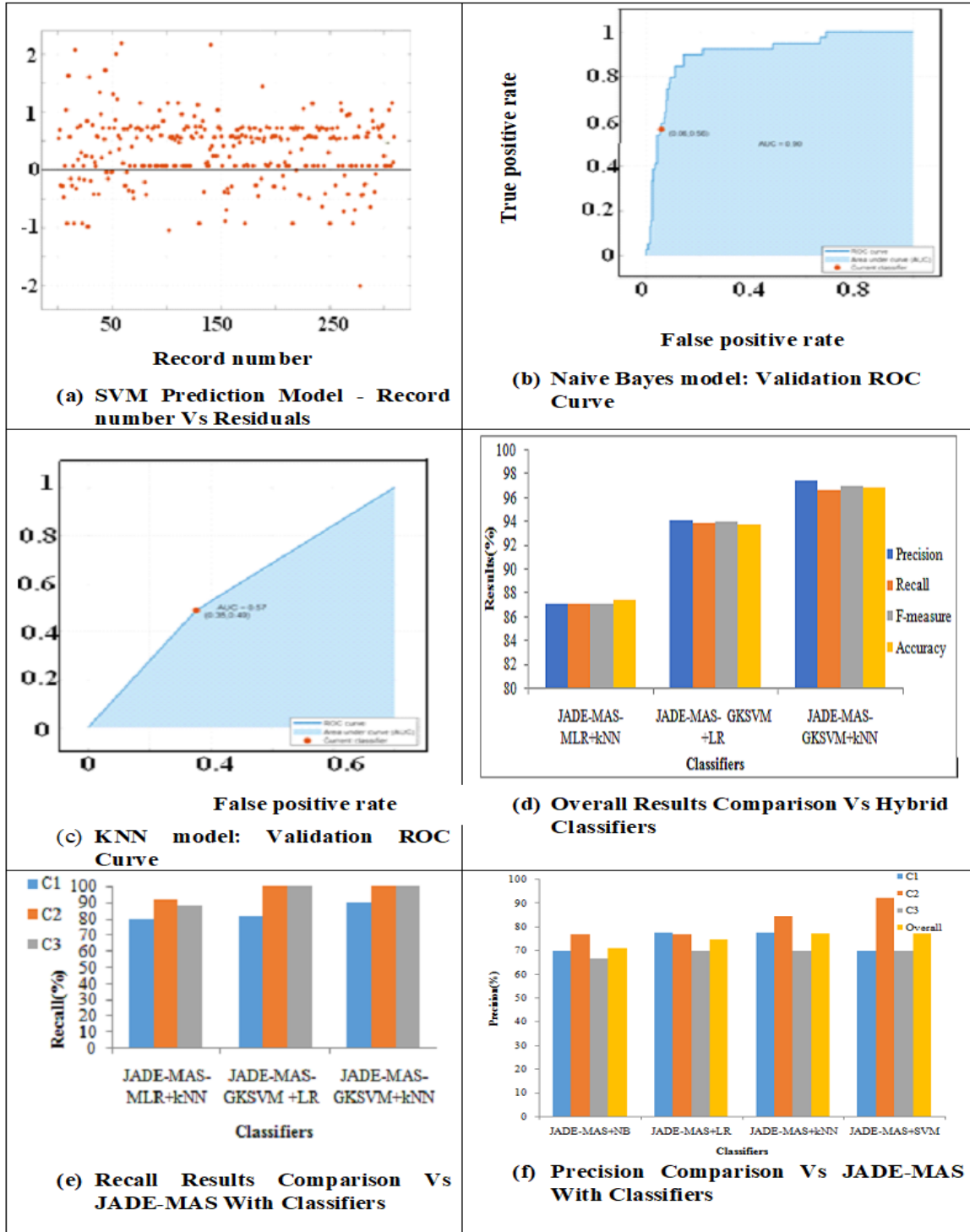
Tab 1 Confusion Matrix of the Lung Cancer Data set Vs Classifiers

METHODS	CLASS	TP	TN	FP	FN	TOTAL
NB	C1	6	19	3	4	32
	C2	10	16	3	3	32
	C3	6	19	4	3	32
LR	C1s	7	19	2	4	32
	C2	10	16	3	3	32
	C3	6	20	4	2	32
KNN	C1	7	20	2	3	32
	C2	11	16	2	3	32
	C3	6	20	4	2	32
SVM	C1	7	20	2	3	32
	C2	11	17	2	2	32
	C3	7	20	3	2	32

Tab 2 Evaluation Metrics Vs JADE-MAS +Classifiers

METHODS	Class	Precision (%)	Recall (%)	F-measure (%)	Accuracy (%)
JADE-MAS +NB	C1	70.00	70.00	70.00	81.25
	C2	76.92	83.33	84.38	84.38
	C3	66.67	60.00	63.16	78.13
AVERAGE (%)		**71.20**	**71.11**	**72.51**	**71.88**
JADE-MAS +LR	C1	77.78	70.00	73.68	84.38
	C2	76.92	83.33	80.00	84.38
	C3	70.00	70.00	70.00	81.25
AVERAGE (%)		**74.90**	**74.44**	**74.56**	**75.00**
JADE-MAS+ KNN	C1	77.78	70.00	73.68	84.38
	C2	84.62	84.62	84.62	87.50
	C3	70.00	77.78	73.68	84.38
AVERAGE (%)		**77.47**	**77.47**	**77.33**	**78.13**
JADE-MAS +SVM	C1	70.00	77.78	73.68	84.38
	C2	92.31	85.71	88.89	90.63
	C3	70.00	87.50	77.78	87.50
AVERAGE (%)		**77.44**	**83.66**	**80.12**	**81.25**

Figure 2 shows the outputs of KNN, SVM, Naïve Bayes and comparison results of JADE MAS

(a) SVM Prediction Model - Record number Vs Residuals

(b) Naive Bayes model: Validation ROC Curve

(c) KNN   model: Validation   ROC Curve

(d) Overall Results Comparison Vs Hybrid Classifiers

(e) Recall Results Comparison Vs JADE-MAS With Classifiers

(f) Precision Comparison Vs JADE-MAS With Classifiers

Fig 2 the outputs of KNN, SVM, Naïve Bayes and comparison results of JADE MAS

23

 The precision results of four classifiers, JADE-MAS+NB, JADE-MAS+LR, JADE-MAS+KNN, and JADE-MAS+SVM, are compared in Figure 2. According to Table 2, the precision scores obtained by these classifiers are 71.20%, 74.90%, 77.47%, and 77.44%, respectively. JADE-MAS+KNN and JADE-MAS+SVM classifiers exhibit higher precision results of ~78% compared to the other two classifiers. The Recall results of the same four classifiers are compared in Figure 2(e). The recall scores achieved by these classifiers are 71.11%, 74.44%, 77.47%, and 83.66%, respectively, according to Table 2. Among all the classifiers, JADE-MAS+SVM performs the best in terms of recall. The F-Measure results of the same four classifiers are shown in Figure 2, and the results obtained by JADE-MAS+NB, JADE-MAS+LR, JADE-MAS+KNN, and JADE-MAS+SVM are 72.51%, 74.56%, 77.33%, and 80.12%, respectively. The proposed JADE-MAS+SVM classifier provides higher F-Measure results than the other methods. **Table 3** presents the experimental results of different classifiers on the Hungarian data set in terms of evaluation metrics such as precision, recall, F-measure, and accuracy.

Tab 3 Performance Comparison Analysis of Hungarian Heart Disease Data set
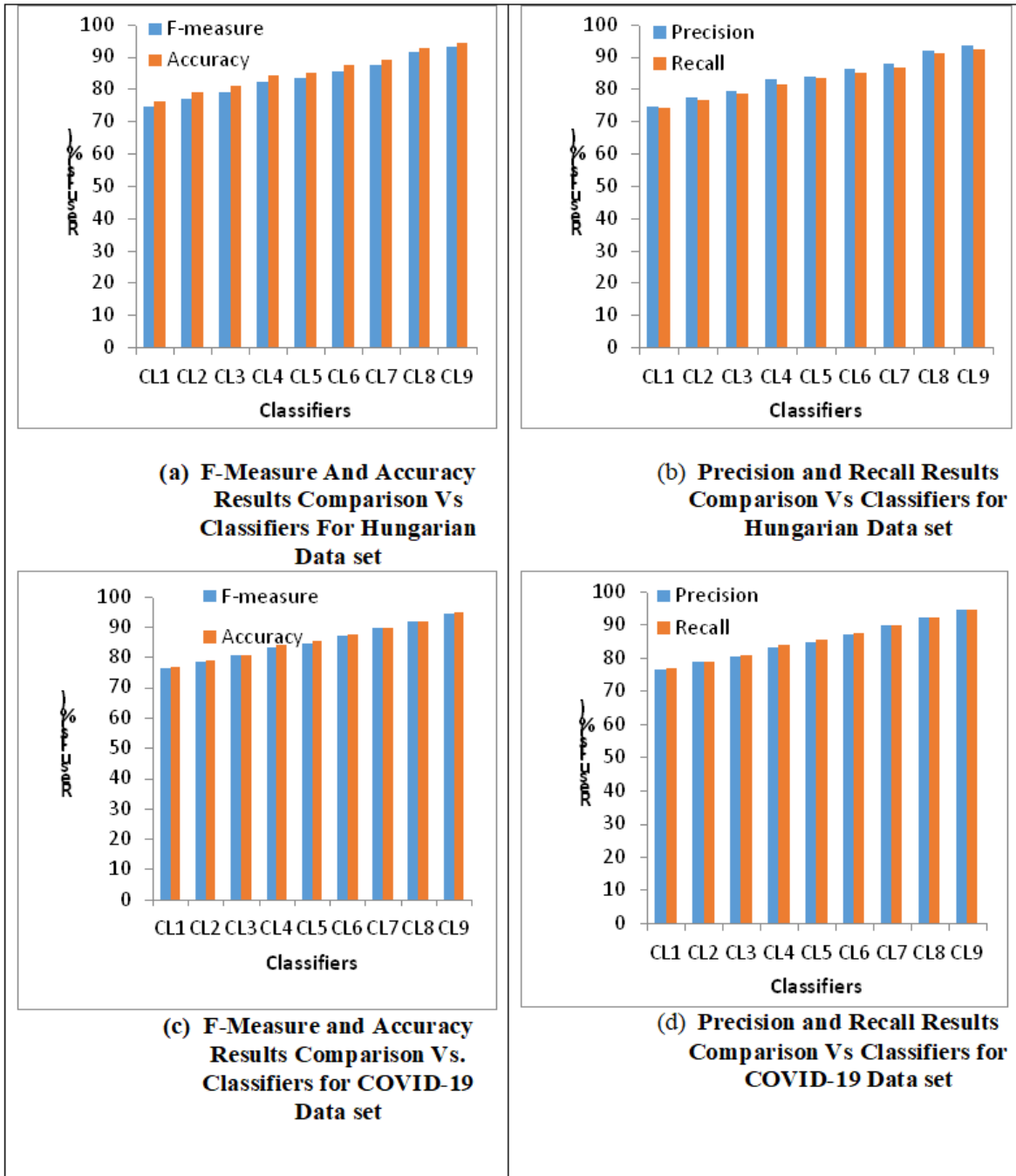
Hungarian data set- Results (%)				
Classifiers	**Precision**	**Recall**	**F-measure**	**Accuracy**
LR	74.70	74.23	74.47	76.19
KNN	77.27	76.57	76.92	78.91
SVM	79.34	78.58	78.96	80.95
JADE-MAS +LR	82.97	81.24	82.09	84.35
JADE-MAS+ KNN	83.63	83.42	83.52	85.03
JADE-MAS + SVM	86.12	85.07	85.59	87.41
JADE-MAS- MLR+KNN	88.01	86.61	87.31	89.12
JADE-MAS- GKSVM +LR	91.87	90.97	91.42	92.86
JADE-MAS- GKSVM+KNN	93.30	92.45	92.88	94.22

The experimental results demonstrate that the JADE-MAS-GKSVM+KNN classifier outperformed all other classifiers in terms of performance assessment measures. The JADE-MAS-GKSVM+KNN classifier achieved 93.3% Precision, 92.45% Recall, 92.88% F-measure, and 94.22% accuracy. The JADE-MAS-GKSVM+LR classifier achieved the second-best result, with 91.87% Precision, 90.97% Recall, 91.42% F-measure, and an accuracy of 92.86%. The JADE-MAS-MLR+KNN classifier yielded 88.01% Precision, 86.61% Recall, 87.31% F-measure, and an accuracy of 89.12%. In contrast, LR's performance was poor, with values of 74.7% Precision, 74.23% Recall, 74.47% F-measure, and 76.17% accuracy. The JADE-MAS-GKSVM+KNN classifier outperformed all other classifiers in terms of precisions, recalls, F-measures, and accuracies. **Table 4** presents the experimental results of the various classifiers on the COVID-19 symptoms dataset with respect to evaluation metrics precision, recall, F-measure, and accuracy.
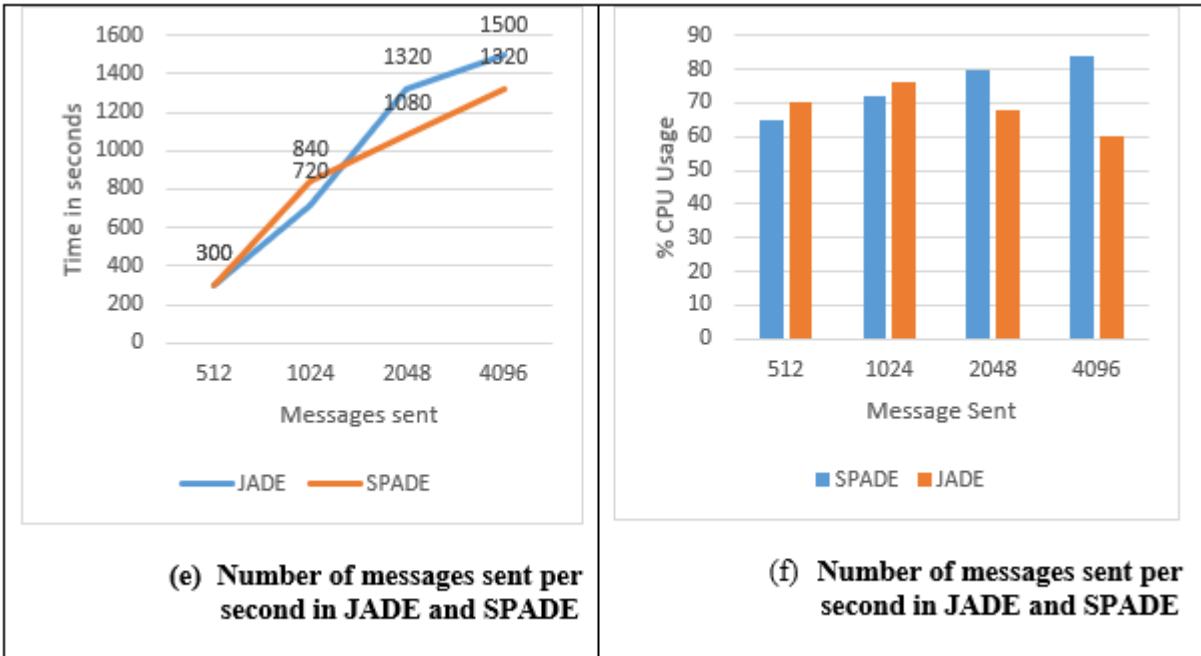
Tab 4 Performance Comparison Analysis of COVID-19 symptoms Data set

COVID-19 symptoms data set- Results (%)				
Classifiers	**Precision**	**Recall**	**F-measure**	**Accuracy**
LR	76.19	76.88	76.08	76.14
KNN	78.59	78.81	78.62	78.59
SVM	80.37	80.57	80.39	80.35
JADE-MAS +LR	83.13	83.77	83.32	83.16
JADE-MAS+ KNN	84.61	85.27	84.68	84.56
JADE-MAS + SVM	87.04	87.37	87.00	87.02
JADE-MAS- MLR+KNN	89.48	89.62	89.51	89.47
JADE-MAS- GKSVM +LR	91.86	91.94	91.86	91.93
JADE-MAS- GKSVM+KNN	94.39	94.47	94.42	94.38

The JADE-MAS-GKSVM+KNN classifier outperformed the other classifiers in terms of all performance evaluation metrics, achieving 94.39% Precision, 94.47% Recall, 94.42% F-measure, and 94.38% accuracy. The JADE-MAS-GKSVM+LR classifier obtained the second-best result, with 91.86% Precision, 91.94% Recall, 91.86% F-measure, and accuracy of 91.93%. The JADE-MAS-MLR+KNN classifier had a Precision of 89.48%, Recall of 89.62%, F-measure of 89.51%, and accuracy of 89.47%. However, the LR classifier had the worst performance, with a Precision of 76.19%, Recall of 76.88%, F-measure of 76.08%, and accuracy of 76.14%. Overall, the JADE-MAS-GKSVM+KNN classifier's predicted outcomes were superior to those of the other classifiers. **Figure 3** displays the JADE and SPADE results.

(a) F-Measure And Accuracy Results Comparison Vs Classifiers For Hungarian Data set

(b) Precision and Recall Results Comparison Vs Classifiers for Hungarian Data set

(c) F-Measure and Accuracy Results Comparison Vs. Classifiers for COVID-19 Data set

(d) Precision and Recall Results Comparison Vs Classifiers for COVID-19 Data set

(e) Number of messages sent per second in JADE and SPADE

(f) Number of messages sent per second in JADE and SPADE

Fig 3 Results of JADE and SPADE

The results demonstrate that the final GKSVM+KNN classifier outperforms the other classifiers in terms of precision and F-measure with values of 94.39% and 94.42%, respectively. In contrast, the other classifiers achieved precision values ranging from 76.19% to 91.86% and F-measure values ranging from 76.08% to 91.86%. Table 5 presents a performance comparison of different algorithms. This section focuses on the performance evaluation of agent systems on the SPADE and JADE platforms after considering feedback from other agents. SPADE agent system proves to be more efficient in CPU utilization for communication compared to JADE. While both platforms experienced increased CPU consumption, JADE's CPU utilization was noticeably lower than SPADE's as the volume of messages transferred increased.

Table 5 Performance Comparison of Different Classifiers

Methods	Accuracy %
NB	68.75
LR	71.88
KNN	75.00
SVM	78.125
JADE-MAS+NB	71.88
JADE-MAS+LR	75.00
JADE-MAS+KNN	78.12
JADE-MAS+SVM	81.25
JADE-MAS-MLR+KNN	87.5
JADE-MAS-GKSVM+LR	93.75
JADE-MAS-GKSVM+KNN	96.88

26

The table above demonstrates that the JADE-MAS with final GKSVM+KNN classifier provides higher accuracy compared to other methods. To enhance the diagnosis of various disorders, it would be more suitable to develop a decision support system using machine learning techniques. Furthermore, some unnecessary features slowed down the performance of the diagnosis system and increased computation time. Hence, an innovative aspect of this study is the use of feature selection algorithms to identify the most suitable characteristics that enhance classification accuracy and reduce execution time. Future research should focus on conducting additional experiments utilizing feature selection methods and optimization techniques to further enhance the predictive classifiers' effectiveness in disease diagnosis. Additionally, another interesting area of research is the development of a multi-agent for the decision-making system using SPADE as a platform.

CONCLUSION

Early and accurate identification of illnesses like lung cancer and heart failure is crucial for treatment and prevention. Traditional medical history approaches to diagnosis have been criticized for being unreliable. Machine learning (ML) techniques have been used to diagnose various diseases, and this study focuses on developing ML classifiers like LR, KNN, NB, and SVM to detect illnesses in medical data sets. Multi-disease categorization using the JADE-MAS approach with hybrid classifiers is introduced, including MLR+KNN, GKSVM+LR, and GKSVM+KNN. Feature selection algorithms are used to identify the best characteristics for increased accuracy and faster execution time. Further testing is necessary to improve the efficacy of these predictive classifiers, and the development of a multi-agent DM system using SPADE as a platform is an area of interest.

REFERENCES

[1] Xiang C, Xiao Y, Qu P, Qu X., 2014 Feb. Network Intrusion Detection Based on PSO-SVM, *Indonesian Journal of Electrical Engineering and Computer Science*, 1;12(2):1502-8.

[2] Zuo WL, Wang ZY, Liu T, Chen HL., 2013 Jul. Effective detection of Parkinson's disease using an adaptive fuzzy k-nearest neighbor approach. *Biomedical Signal Processing and Control*, 1; 8(4):364-73.

[3] Wan CH, Lee LH, Rajkumar R, Isa D., 2012 Nov, A hybrid text classification approach with low dependency on parameter by integrating K-nearest neighbor and support vector machine, *Expert Systems with Applications*, 1;39(15):11880-8.

[4] Tolles, J. and Meurer, W.J., 2016. Logistic regression: relating patient characteristics to outcomes. *Jama*, *316*(5), pp.533-534.

[5] Malik A. and J. Iqbal, 2016. Extreme learning machine based approach for diagnosis and analysis of breast cancer, *Journal of the Chinese Institute of Engineers*, vol. 39, no. 1, pp. 74–78.

[6] Chang, C.L. and Chen, C.H., 2009. Applying decision tree and neural network to increase quality of dermatologic diagnosis. Expert Systems with Applications, 36(2), pp.4035-4041.

[7] Chauhan D. and V. Jaiswal, 2016. An efficient data mining classification approach for detecting lung cancer disease, *International Conference in Communication and Electronics Systems (ICCES)*, pp. 1–8.

[8] Souri, E.A., and Tsoka, S., 2020. Cancer Classification through a Hybrid Machine Learning Approach, *The Twelfth International Conference on Bioinformatics, Biocomputational systems and Biotechnologies*, pp. 20-27.

[9] Atteya, W.A., Dahal, K. and Hossain, M.A., 2010. Multi-agent system for early prediction of urinary bladder inflammation disease. *In 10th International Conference on Intelligent Systems Design and Applications*, pp. 539-544.

[10] Uppal, K. and Lee, E.K., 2017. optSelect: using agent-based modeling and binary PSO techniques for ensemble feature selection and stability assessment. bioRxiv, pp.1-11.

[11] Pourpanah, F., Tan, C.J., Lim, C.P. and Mohamad-Saleh, J., 2017. A Q-learning-based multi-agent system for data classification. *Applied Soft Computing*, 52,pp. 519-531.

[12] Kavak, H., Padilla, J.J., Lynch, C.J. and Diallo, S.Y., 2018. Big data, agents, and machine learning: towards a data-driven agent-based modeling approach. *In Proceedings of the Annual Simulation Symposium*, pp. 1-12.

[13] Kuttikrishnan, M., 2010. A novel approach for cardiac disease prediction and classification using intelligent agents. *arXiv pre-print* arXiv:1009.5346, pp.1-9.

[14] Zhao, H., Xu, W., Li, L. and Zhang, J., 2011. Classification of breast masses based on multi-view information fusion using multi-agent method. *In 5th International Conference on Bioinformatics and Biomedical Engineering*, pp. 1-4.

[15] Tatari, F., Akbarzadeh-T, M.R. and Sabahi, A., 2012. Fuzzy-probabilistic multi agent system for breast cancer risk assessment and insurance premium assignment, *Journal of Biomedical informatics*, 45(6), pp. 1021-1034.