JOURNAL OF SOFT COMPUTING AND DATA MINING
e-ISSN: 2716-621X

JSCDM

Vol. 6 No. 1 (2025) 230-246
https://publisher.uthm.edu.my/ojs/index.php/jscdm

An Ensemble Learning Model for Multi-Type Cancer Prediction in Clinical Diagnostic Decision Support Systems

Salama A Mostafa[1]*, Dhafar Fakhry Hasan[2], Maha A. Abdul-Jabar[2], Abdelrahman H. Hussein[3], Aida Mustapha[4]

[1] *Department of Artificial Intelligence Engineering Techniques, College of Technical Engineering, Alnoor University, Mosul 41012, Nineveh, IRAQ*
[2] *Computer Unit, College of Medicine, University of Mosul, Mosul 41012, Nineveh, IRAQ*
[3] *Networks and Cybersecurity Department, Hourani Center for Applied Scientific Research, Al-Ahliyya Amman University, Amman, JORDAN*
[4] *Centre for Artificial Intelligence and Data Science, Universiti Malaysia Pahang Al-Sultan Abdullah, 26300 Kuantan, Pahang, MALAYSIA*

*Corresponding Author: salama.adress@alnoor.edu.iq
DOI: https://doi.org/10.30880/jscdm.2025.06.01.015

Article Info

Received: 4 May 2025
Accepted: 25 June 2025
Available online: 30 June 2025

Keywords

Machine learning (ML), ensemble learning (EL), ensemble learning, Decision Support System (DSS), multi-cancer diagnosis, breast cancer, liver cancer, cervical cancer, brain cancer

Abstract

Different organs of the human body are targeted by cancer, leading to varied effects that cause numerous severe health conditions. The early detection of cancer is essential, as it provides timely clinical admission of the patient, which is necessary for successful treatment. Machine Learning (ML) and Deep Learning (DL) algorithms are being widely used to detect and identify cancer cases, relying on data from multiple disciplines, including medical, biomedical, and bioinformatics. These algorithms can identify meaningful diagnostic patterns and detect cancer cases in complex cancer databases. In this paper, we have suggested an ensemble learning approach that consists of five common ML and DL algorithms to build an ensemble-based Multi-Type Cancer Prediction (eMTCP) model. The algorithms used in the development of the eMTCP model are Naive Bayes (NB), Random Forest (RF), Support Vector Machines (SVM), Convolutional Neural Networks (CNNs), and Long Short-Term Memory (LSTM). The eMTCP is utilized in the development of a Cancer Diagnostic Clinical Decision Support System (CDCDSS). Four cancer diagnostic datasets, namely liver cancer, breast cancer, brain cancer, and cervical cancer, are then used to compare the performance of the eMTCP model. The algorithm with the best potential is eMTCP (stacked ensemble), which yields the best F1-scores across all datasets: breast (0.979), liver (0.765), brain (0.898), and cervical (0.482), demonstrating superior performance in multi-type cancer prediction. CNN and LSTM achieved high stability with superior F1-scores of 0.957 (breast), 0.669 (liver), and 0.853 (brain). CNN outperformed in the case of liver cancer and showed similar performance in other cancer types. The SVM ML model has the lowest scores of all: 0.969 (breast), 0.716 (liver), 0.844 (brain), and only 0.153 (cervical), indicating that its representation in the clinical-only dataset is not very reliable.

1. Introduction

Worldwide, cancer continues to be a leading cause of death, resulting in 9.7 million fatalities from 20 million cases in 2022, accounting for 16.44% of global deaths as reviewed by the World Health Organization (WHO) in https://www.who.int/news-room/fact-sheets/detail/cancer. The global cancer landscape deteriorates as populations age rapidly and environmental hazards increase due to behavioral changes. New cancer diagnoses worldwide primarily involve breast, lung, colorectal, prostate, and stomach cancers, with breast cancer topping the statistics at 2.26 million new diagnoses [1]. Despite significant advancements in screening systems and treatment options, doctors still face challenges in detecting cancers and providing equal diagnostic tools to all individuals, especially in impoverished and developing nations. Over the years, there has been a persistent effort toward automating cancer diagnosis research [2]. In this context, various scientific methods have been employed to identify cancer cases at an early stage. Additionally, techniques also exist for predicting treatment outcomes in advance. This has been greatly supported by the vast amounts of data generated by modern technologies, including artificial intelligence, the Internet of Medical Things, virtual reality, and digital twins, among others [3], [4].

In the field of Artificial Intelligence (AI), Machine Learning (ML) algorithms effectively map medical data samples to class labels [2], [5] by (a) approximating unseen dependencies in a given dataset and (b) predicting new outputs within a system in the medical informatics domain [6], [7]. In the first approach, a set of labeled data is used for training to map input and output data. Conversely, the second approach does not involve any labeled data, as it seeks to find patterns in input data rather than systematic input-output mapping during the training process. In supervised learning, classifying objects is a process that categorizes input data into a set of classes [8], [9]. Generally, ML algorithms, including Artificial Neural Networks (ANN), Naive Bayes (NB), Random Forest (RF), Decision Trees (DT), and Support Vector Machines (SVM), involve processing various types of data, extract their features, analyzing their characteristics, and making decisions that assist computer systems and applications [10]-[15]. These algorithms have been applied to numerous healthcare applications, aiding in disease diagnosis, medical image analysis, Clinical Decision Support Systems (CDSS), patient risk assessment, mental health assessment, and other computer-aided diagnosis systems (CADS) [16]-[19].

On the other hand, deep learning (DL) has enhanced both the development and performance of AI-based CAD systems, particularly in cancer detection and diagnosis. DL has introduced complex artificial neural network (ANN) algorithms, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) [20], [21]. These algorithms automatically extract medical image features, genomic data, and information from electronic health records, eliminating the need for manual effort from experts. This allows doctors to diagnose more accurately, view images more quickly, and detect breast, lung, prostate, and colorectal cancers at the earliest possible stage [22], [23]. With the aid of DL, CADS can identify early indications in mammography (X-ray), Computed Tomography (CT) scans [24], and Magnetic Resonance Imaging (MRI) [25], as well as in histopathology, resulting in fewer false-positive cases. They also enable the identification of risk groups, the development of treatment plans, and the prediction of outcomes using various medical data. For this reason, DL represents a critical technology for creating intelligent, scalable, and accurate diagnostic tools that support clinicians in oncology [22], [24].

Eventually, ML and DL methods have garnered significant attention from the relevant research community as a means of diagnosing cancer through the introduction of various CDSSs. However, enhancing the prediction accuracy for cancer in complex conditions remains one of the major challenges, including cancer heterogeneity across subtypes and types, unbalanced dataset distributions, and opaque diagnosis models that undermine clinical confidence and prediction reliability. Prediction accuracy is further compromised by medical records that may contain noise and incompleteness, as well as the variability encountered in data acquisition methods [4], [6]. Additionally, these algorithms have a limited scope and specialization, in which most systems are designed to detect or support decisions for a single type of cancer, which restricts the system's ability to analyze broader patterns, recognize interrelated features across cancer types, or handle cases where symptoms and biomarkers may overlap [26]- [28]. As a result, such systems may lead to incomplete diagnostics, misclassification, or the failure to detect secondary malignancies.

An important aspect of the modern profession in soft computing research within ML and DL is the development of automated CADS in which ML or DL algorithms enhance disease diagnosis decisions by automatically constructing a mathematical model from sample data during the training phase and predicting the diagnostic outcomes of new cases [22]. Subsequently, combined ML and DL algorithms, such as in ensembles, feature- or decision-level fusion, transfer learning, and hybrid learning approaches, can identify patterns and relationships within large datasets, making them highly effective at predicting multiple cancer types [5], [26]. A multi-type cancer diagnosis approach addresses this limitation by enabling the CDSS to analyze heterogeneous data from various cancer types, identify shared and distinct diagnostic features, and offer more holistic and inclusive decision support. It enhances the system's diagnostic accuracy, generalizability, and clinical usefulness, especially in complex or ambiguous cases where multiple cancer types may be present or suspected [29], [30].

Making a CDSS for multi-type cancer diagnosis is necessary since it helps to analyze a broader range of cancer-related information from many organs and body systems. Because it compares several cancer types, a multi-type system can spot common biomarkers, the same or related symptoms, and similar risks, which boosts the precision and detail in cancer diagnosis. when the system reviews data for multiple cancers, it understands disease signs better, which results in clearer distinctions between similar diseases. Taking into account a variety of factors helps professionals diagnose more accurately and uncover multicultural cancers, which results in making more informed treatment decisions. Defining cancer diseases through various classifications enhances the accuracy and adaptability of cancer prediction models, allowing medical professionals to analyze real-world patient data.

This research aims to propose an ensemble multi-type cancer prediction model for developing a cancer diagnostic clinical DSS. The ensemble learning model comprises five widely used ML and DL models: NB, RF, SVM, RNN, and LSTM. The main contributions of this research are as follows:

- Initially, five ML algorithms: NB, RF, SVM, CNN, and LSTM have been developed to assess their ability to classify multi-type cancers. This part contributes benchmark results of the ML and DL algorithms across four medical datasets concerning breast cancer, cervical cancer, brain cancer, and liver cancer.
- The architectural design of an Ensemble-based Multi-Type Cancer Prediction (eMTCP) model includes five ML and DL algorithms and a two-level heterogeneous stacking aimed at enhancing the classification of multi-type cancers.
- The deployment of the eMTCP for developing a Cancer Diagnostic Clinical Decision Support System (CDCDSS). The CDCDSS is tested to assess its ability to classify multiple types of cancers.

The structure of this paper is organized into five sections, with Section 2 detailing the related work and discussing the role of ML and DL algorithms in cancer diagnosis. The materials and methods used in this study are outlined in Section 3, which covers the research methodology for testing and evaluating the contributions of this research. Section 4 presents the results obtained and discusses the research findings. Lastly, the conclusion and future work are effectively summarized in the final section.

2. Related Work

This review aims to provide insight into key ML and DL methods commonly used for classification problems and their relevant applications in modeling automated cancer diagnosis. It discusses research papers that employ various algorithms for cancer diagnosis, including liver cancer [2], cervical cancer [4], breast cancer [6], and brain cancer [7].

Yadav et al. [6] compared various algorithms, specifically ANN, DT, RF, NB, SVM, and K-NN. The authors trained these models using a dataset containing parameters describing patient tumor conditions for breast cancer. Subsequently, the trained model was employed to identify patients with such tumors. The results indicate that SVM and RF achieve the highest accuracy of 97.2%. Furthermore, NB has the highest precision and recall rates of 97.2% and 97.1%, respectively. Omondiagbe et al. [8] examined the performance of SVM (using a radial basis kernel), ANN, and NB with the Wisconsin Diagnostic Breast Cancer (WDBC) dataset. The maximum accuracy obtained was 98.82%, with a sensitivity of 98.41%, a specificity of 99.07%, and an area under the curve (AUC) of 99.94%. Cherif [9] introduced a new approach to enhancing the K-NN algorithm, aiming to boost its performance by accelerating the process. Consequently, the research contributed to clustering data samples, selecting the most suitable features, and measuring similarities using reliability coefficients. The study demonstrated that this new solution outperformed commonly used classification approaches, achieving a 94% average F-measure on the selected dataset.

A similar study by Kharya et al. [11] implemented a breast cancer predictive system based on the NB classifier with an accuracy of 93 percent. Moreover, the authors designed a graphical user interface (GUI) for handling patient records. In this model, patients with breast cancer are allowed to predict their likelihood of experiencing symptoms. Mostafa et al. [12] achieve their study goals by focusing on the conditions under which diverse feature assessment and classification techniques can be used to enhance accuracy in the diagnostic process of Parkinson's disease. This study aims to identify the key characteristics of biomedical data. It explores different ML classifiers that are SVM, K-NN, and DT, on a performance basis, where the performance result is checked through accuracy, sensitivity, and specificity. The paper suggests that the right features and classifier could be of great use in improving the accuracy of Parkinson's disease diagnosis and in the potential of intelligent systems to provide early and accurate diagnoses thereof.

In the study conducted by Obaid et al. [13], the authors concentrate on determining breast cancer through several ML methods applied to the Wisconsin Breast Cancer Dataset analysis. The researchers in this study compare the accuracy and the precision of the DT, SVM, and NB algorithms. The results indicate that certain models outperform others in distinguishing between benign and malignant tumors. This proves the assumption that the choice of classification techniques is a key factor in performing diagnosis and shows that ML has the potential to be used in creating efficient tools that promote the diagnosis of breast cancer.

Priya et al. [14] also investigated the performance of ML algorithms in predicting the outcomes of Indian liver patients, whose datasets are unbalanced, consisting of a collection of samples with 72 percent of the samples belonging to liver patients and 28 percent to non-liver patients. The findings indicate that the DT (J48) algorithm has the best accuracy level of 95.04%. Durai et al. [15] used a few ML algorithms on data pertaining to liver disease patients to evaluate the qualities of their classification algorithms in a similar study. The DT (J48) algorithm achieved the highest accuracy of 95.04% after feature selection.

A comparative study on NB, KNN, and SVM algorithms was done by Sagala [16] to diagnose malignant cancer samples. These techniques were used to provide four kinds of medical tests with various target parameters. The results showed that the NB algorithm performed better than others when tested using 10-fold cross-validation. Wahid and Al-Mazini [17] employed a classification algorithm enhanced by ant colony optimization to analyze cervical cancer data, achieving an accuracy of over 90 percent.

Ceylan and Pekel [18] investigated the robustness of multi-label classification techniques for the early-stage diagnosis of cervical cancer. The ML algorithms used include NB, J48 DT, and RF. The authors compared the methods using various metrics, including accuracy, Hamming loss, exact match (subset accuracy), and ranking loss, for performance evaluation. The accuracy percentages for the examined algorithms are approximately above 80%, except for J48 DT, which showed a lower accuracy percentage.

Singh [19] initially employed the Genetic Algorithm (GA) for feature selection, aiming to create a more effective feature set for classification with ML-based classifiers. The classifiers used in this research are SVM, RF, and GBM, along with the oversampling technique SMOTE. Bayesian optimization is applied for hyperparameter tuning to enhance the true positive accuracy of the models mentioned earlier. The results reveal that GBM exhibits the highest sensitivity at 77.8%, followed by SVM with a sensitivity of 55.58% and RF with a sensitivity of 44.4%.

DL is primarily applied to medical images to fully utilize its feature extraction capabilities. Saba et al. [20] presented a DL model designed to identify brain tumors from MRI scans. They employed a two-stage DL process: first, images were segmented using U-Net and ResNet50; second, transfer learning facilitated the classification of tumors with ResNet50-U-Net, DenseNet201, MobileNet V2, Inception V3, and NASNet. They improved model performance by preprocessing and augmenting the data. Among all the tested models, NASNet achieved a remarkably high accuracy of 99.6%, surpassing ResNet50 and DenseNet201. This demonstrates how relying on effective segmentation and powerful transfer learning methods enhances the recognition and classification of brain tumors using MRI.

Kaur et al. [21] recommend utilizing CT scan images to develop a hybrid DL model for predicting liver cancer in their 2022 study. They employ freely available pre-trained CNN models found on the internet, stacking them into a single network, which simplifies the model's ability to comprehend rich image details without requiring complex datasets. When evaluated with biopsy-based CT images, the model achieved outstanding results, scoring 99.5% accuracy, 86.4% precision, and 97.9% recall, significantly outperforming single-model systems. Although it was not possible to use large datasets, the study's findings suggest that combining various DL networks can facilitate early detection of liver cancer and provide clinicians with reliable support.

The 2025 research by Kaddes et al. [23] employed a combination of a CNN and an LSTM to identify breast cancer using images from Kaggle data. CNN helps identify different parts and look of the tumor using spatial features, and the LSTM layer manages the order and ties between these important features. In comparison to standalone models and GRU, VGG-16, and ResNet-50, the CNN-LSTM model performed significantly better, achieving accuracies of 99.17% and 99.90% for the two datasets, and scoring high in terms of sensitivity, specificity, F-score, and AUC. Analysts suggest that combining spatial and sequential data in a joint CNN-LSTM system significantly enhances the accuracy and reliability of breast cancer classification.

An ensemble model enables the combination of clinical and imaging data for multi-type cancer diagnosis. Consolidation of various sources of data (pictures and medical records) increases the accuracy of the model and its performance in recognizing similar patterns across patients with various types of cancer. In turn, Wang et al. [26] produced a multimodal system analyzing images and clinical data using both DL and traditional ML, thus being able to detect several types of cancer. Their approach utilizes convolutional neural networks to recognize spatial imaging features, while XGBoost processes the clinical details. Subsequently, a meta-learner applies stacking or soft voting to combine the earlier results and reach the final diagnosis. Considering three cancer types and various datasets, the ensemble model demonstrated superior accuracy, AUC, and stability compared to models that relied on a single technique. These results demonstrate that integrating images and health records through such strategies significantly enhances the accuracy of cancer detection, providing clinicians with more informative insights.

Ali et al. [29] claim that having a single model that works with multiple cancer types and forms has advantages for learning and enhances the system's reliability. Accordingly, they designed EDRNet, giving special attention to incorporating a deep residual architecture that functions across various tumor and polyp types, as well as several scanning methods, including brain MRI, breast ultrasound, and colorectal polyp endoscopy. The system employs Spatial-Channel Fusion Attention (SCFA) modules, along with enhanced versions of residual dilated blocks, as well as a refined Attention Feature Fusion Module built on an EfficientNet encoder. Trained with a diverse array of cancer datasets, the model improves IoU by 9% for brain tumors, 4% for breast tumors, and 1.5% for colorectal polyps. Its ability to detect both global trends and minute characteristics helps meet the needs for accurate and comprehensive cancer assistance.

In a 2025 study, Wang et al. [30] describe CPLOYO, a modern model designed explicitly for identifying pulmonary nodules, which is vital for diagnosing various types of lung cancer. To enhance YOLOv8, they implement three crucial modules: (1) C2f_RepViTCAMF merges lightweight Vision Transformer blocks with rep-based convolutions and attention to better identify small nodules; (2) MSCAF unifies the capturing of nodules of various sizes; and (3) the KAN module boosts the representation of diverse features. By testing on the LUNA16 dataset, CPLOYO demonstrates significant improvements, with a precision increase of 4.5%, a recall improvement of approximately 7%, and mAP50 gains of around 4.4% compared to YOLOv8, while still operating efficiently in resource-limited medical settings. The stable and accurate detection of tumors in patients with lung cancer is enhanced by CPLOYO, which utilizes both multi-scale and nonlinear ML strategies.

The literature demonstrates that significant progress has been made in developing intelligent cancer diagnosis systems through the use of ML and DL. Numerous studies have demonstrated high accuracy in detecting various types of cancer. Similarly, Mostafa et al. [12] and Obaid et al. [13] proved that selecting the right features and applying ML classifiers enhanced the detection success of Parkinson's disease and breast cancer. Additionally, models that integrate CNN and LSTM [23] and transfer learning [21] have been demonstrated to classify types of cancer effectively. Furthermore, by employing attention mechanisms and integrating features at different scales, EDRNet [29] and CPLOYO [30] have improved tumor segmentation and detection outcomes. Although their findings are promising, they have only been applied to identifying single cancer types, which limits their usefulness for patients with multiple or different cancer types in real-world hospitals. To address this challenge, Wang et al. [30] introduced a method to utilize both images and clinical information collaboratively in diagnosing various types of cancers, demonstrating the effectiveness of joint systems. Currently, few CDSS are equipped to handle multiple cancer types due to the integration of both ML and DL within a single system. This research area is underdeveloped, indicating a need for experts to develop shared models for cancer diagnosis that utilize various data types and intelligent algorithms to support diverse cancer types within healthcare systems.

3. Methods

This methodology aims to develop an Ensemble-based Multi-Type Cancer Prediction (eMTCP) model that utilizes both ML and DL to assist in diagnosing cancer, particularly liver, breast, brain, and cervical cancer types. By enabling the eMTCP model to utilize various data types and innovative combination methods, the development of a Cancer Diagnostic Clinical Decision Support System (CDCDSS) will achieve accurate and transparent performance, effectively classifying different types of cancer. Figure 1 and the subsequent steps outline the phases of the research methodology employed in this work to achieve its objectives.

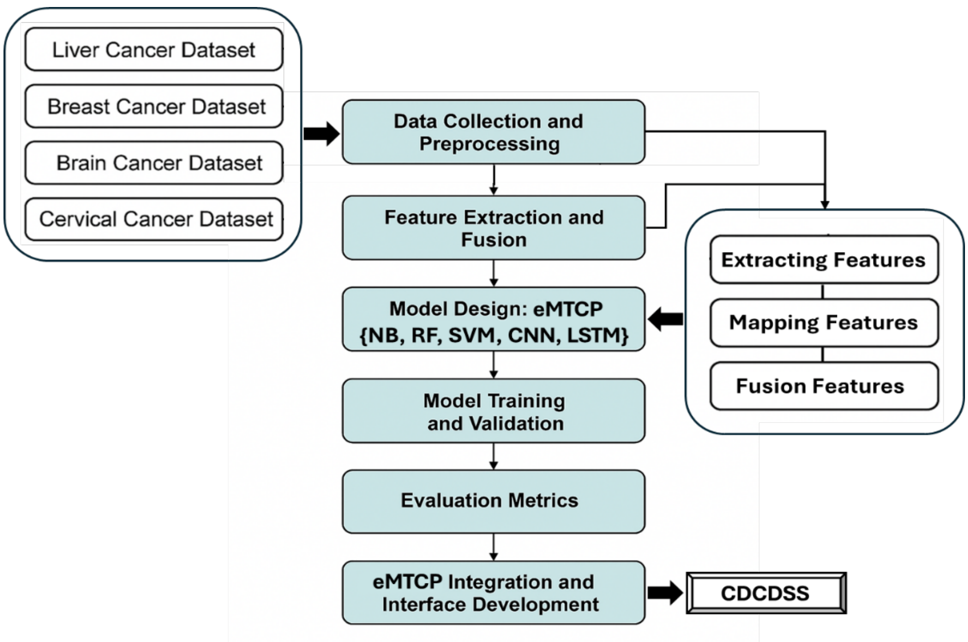

Fig. 1 *The research framework*

1. Data Collection and Preprocessing: Initially, four publicly accessible and reliable datasets related to liver, breast, brain, and cervical cancer were gathered. The data includes clinical information such as blood test results, potential risk factors, personal details, and medical imaging (such as X-Ray, CT scans or MRI). One of the preprocessing steps involves addressing missing data by imputing these values using the mean or the mode. Additionally, the clinical data has been normalized and encoded to facilitate its integration with the image information.
2. Feature Extraction and Fusion: The next step is to identify the important features present in all types of data from medical records. Feature selection is performed based on the importance of features using the SHAP (Shapley Additive Explanations) method for structured clinical data and the importance of regions using Grad-CAM (Gradient-weighted Class Activation Mapping) for medical images. The Principal Component Analysis (PCA) for clinical data and Autoencoders for medical images techniques have been applied to reduce the dimensionality of the datasets and process the data more efficiently. Feature mapping involves aligning the dimensional scales and structures of features into a compatible vector space by applying z-score normalization, allowing the extracted features to be concatenated meaningfully. The datasets are then merged into a multi-type cancer diagnostic dataset by creating a single, multi-modal feature vector and adding the cancer_type feature with values {0: liver, 1: breast, 2: brain, 3: cervical}, which provides a more comprehensive representation of all the cancer cases. To address any class imbalances in the data, both data augmentation and weighting methods are employed. The resulting multi-type cancer diagnostic dataset is well-distributed, ensuring that every type of cancer is represented in each dataset folder.
3. Model Design: A combination of various ML and DL algorithms is employed to develop an eMTCP model. The base models include CNN and LSTM DL algorithms, as well as NB, RF, and SVM ML algorithms, which are used for medical data analysis. All models operate simultaneously, and the results from each model are sent to a meta-classifier that applies the Stacked Generalization (two-level heterogeneous stacking) technique to determine the final outcome. Training as a group is used to help the design perform diagnostics consistently, utilize various types of data, and produce uniform results across different types of cancer.
4. Model Training and Validation: Initially, each ML component of the model is trained and tested separately before being combined into the final eMTCP model, setting a benchmark for each of the five algorithms. The multi-cancer diagnostic dataset is divided into 80% training and 20% testing. This approach provides reliable results and avoids the issue of label distribution across a multivariate dataset. The final eMTCP model is also tested under the same setting.
5. Evaluation Metrics: The performance of the eMTCP model is evaluated using several metrics, including accuracy, precision, recall, F1-score, and AUC-ROC, which are calculated for each type of cancer. Matrices are created to analyze how predictions are made and to identify common classification errors.
6. eMTCP Integration and Interface Development: eMTCP integration refers to combining the methods of the eMTCP model to build the CDCDSS method, while Interface Development focuses on building the connection between them and the end user. Additionally, SHAP and Grad-CAM have been employed to help clinicians visualize and interpret how the model arrives at its decisions, thereby fostering trust and transparency. All different functions are ultimately accessed through the interactive mode of the CDCDSS. Using PyQt5, TensorFlow, and PyTorch frameworks to develop a desktop version of the CDCDSS enables clinicians to enter patient information and receive instant diagnostics. The system presents predictions, the probability of risk score of each result, and explanations in the form of charts outlining the aspects of a patient's clinical information that influenced the result.

3.1 Multi-Type Cancer Prediction

Breast cancer in women is the most frequently diagnosed life-threatening disease and the leading cause of cancer-related death among women. Breast cancer may cause a variety of symptoms, which can appear differently in patients. To improve breast cancer outcomes in areas with inadequate health systems where most women are diagnosed at late stages, it is crucial to emphasize early diagnosis programs that focus on recognizing early signs and symptoms, along with timely diagnosis and treatment referrals [4]. Among women, cervical cancer is the fourth most common type of cancer, with 570,000 new cases reported in 2018. Therefore, it is essential to reduce the high global mortality rate through a multi-pronged strategy involving early diagnosis, prevention, screening, and treatment efforts. However, the disease is often not diagnosed in many low-resource settings until it is more advanced or until available treatments become ineffective, resulting in higher death rates [3].

Another type of cancer is brain cancer, which presents a variety of symptoms, including seizures, sleepiness, confusion, and other changes in behavior. Note that not every brain tumor is cancerous; benign tumors may also produce similar symptoms to malignant tumors. Furthermore, such patients may experience other medical issues; therefore, they require regular tests such as liver function tests, blood analysis, and electrolyte tests, among others. Lastly, the discussion concerns liver cancer diagnosis and treatment. It is observed that symptoms are usually not apparent in the early stages; therefore, it is often diagnosed at later, more advanced stages. In addition to swelling and bleeding, liver cancer may also cause swollen veins that are noticeable under the abdominal skin [5].

The biological features of liver cancer intersect with other multi-type cancers, such as breast cancer, brain cancer, and cervical cancer, which create diagnostic challenges and various patterns of disease evolution, making collective detection methods extremely difficult. Accurate diagnosis of liver cancer happens late since the disease progresses silently, but brain tumors require comprehensive imaging and histopathology testing for proper classification, according to Kaur et al. [21]. Detecting breast and cervical cancers during routine screening is crucial for assisting millions of women globally, as these cancers respond favorably to early diagnosis. A multi-cancer diagnostic framework must integrate clinical and imaging information alongside various molecular observations due to the distinct diagnostic situations. Researchers develop ML ensemble models that simultaneously identify multiple cancer types to create holistic clinical decision systems that enhance patient outcomes while aiding oncologists in treatment planning [28], [20].

In recent years, ML and DL have been among the most intriguing approaches used for building automated systems that help in cancer detection and classification in the field of oncology. However, as overlaps in clinical symptoms, biomarkers, and even the costs of unified models emerge, interest in multi-type cancer prediction is increasing. Multi-cancer diagnostic systems treat the diagnosis problem of various cancer types (e.g., liver, breast, brain, cervical) with multi-omics, medical imaging, or clinical records. Additionally, it holds promise for the early detection of many cancers in these systems and the application of this technology in personalized medicine, a comprehensive approach.

Recent examples of multi-type cancer prediction include the work of Chen et al. [28], who proposed a DL framework using a fully connected neural network to classify 10 different cancer types from The Cancer Genome Atlas (TCGA) gene expression data. They demonstrated strong accuracy and stability, as their model among the classes obtained an F1 score greater than 0.95. Several obstacles to the recognition of histologically similar cancers, due to data imbalance and biological overlap, were also noted in the study. Wang et al. [26] applied a combination of CNNs with gradient boosting machines into a hybrid ensemble model, classifying liver, breast and brain cancers based on a multimodal data including any imaging data and any additional clinical data (imaging and clinical). The findings suggested that the model is moderately accurate (93.6%). With the application of SHAP values, it became more interpretable, allowing for awareness of noteworthy predictive variables. In the meantime, Zhou et al. [27] introduced a transformer-based architecture that applies the attention mechanism to predict cancer type using histopathological images. They tried their model on a dataset of images of breast, liver, cervical, and brain tumors. The study demonstrated the potential of attention-based techniques in modeling complex spatial patterns in cancer diagnosis, particularly in terms of superior performance compared to standard CNN models.

3.2 Datasets

Cancer datasets play a crucial role in the training and validation of ML and DL models for cancer diagnosis, prediction, and prognosis. Such datasets commonly comprise imaging data (e.g., MRI, CT, histopathology) and/or clinical records (e.g., biomarker levels, demographics, medical history). With the help of publicly available and curated datasets, researchers may create a reliable and generalizable CDCDSS. In this study, we selected four representative cancer datasets as presented in Table 1, each belonging to a distinct organ system and data modality as follows.

- Cervical Cancer Dataset: This clinical data includes patient-level data of 858 women with demographic, lifestyle, and clinical history data, such as smoking, number of sexual partners, STDs, use of hormonal contraceptives, and Pap smear outcomes. It consists of 36 binary or numerical valued attributes and four target variables associated with the diagnoses of cervical cancer (Hinselmann, Schiller, Cytology, and Biopsy). The data is helpful in risk prediction models, which take in structured tabular data.
- Breast Cancer Dataset: The Curated Breast Imaging Subset of DDSM (CBIS-DDSM) is a labeled subset of the Digital Database for Screening Mammography (DDSM). It offers mammographic images and ROI masks, BI-RADS recommendations, pathology-defined diagnoses, and density categories. It contains 2,620 DICOM format mammograms, which can be utilized in DL-based mass detection, classification (between benign and malignant), and image segmentation.
- Liver Cancer Dataset: LiTS (Liver Tumor Segmentation Challenge) is a benchmark dataset and contest meant to improve automatic liver and liver tumor segmentation algorithms based upon CT (Computed Tomography) scans. It is a very common and widely authoritative dataset in medical image processing, specifically for liver cancer detection and delineation.

- Brain Cancer Dataset: The Figshare Brain MRI Brain Cancer dataset comprises brain MRI images, categorized into four classes: glioma tumor, meningioma tumor, pituitary tumor, and no tumor. It consists of 3264 T1-weighted contrast-enhanced images annotated by professional radiologists. It has become quite popular in CNNs and transfer learning models on brain tumor classification and segmentation. The data is organized into folders by category, making it easy to use in supervised learning.

Table 1 *The sources of the multi-type cancer diagnostic dataset*

Cancer Type	Dataset Name	Modality	Task	Source
Cervical Cancer	UCI Risk Factors	Tabular Clinical data	Risk Prediction	UCI
Breast Cancer	CBIS-DDSM	Imaging + Clinical	Mass Detection, Diagnosis	TCIA
Liver Cancer	LiTS	Imaging (CT)	Segmentation, Detection	LiTS
Brain Cancer	Figshare MRI	Imaging (MRI)	Tumor Classification	Figshare

3.3 The eMTCP Model

This paper presents a model of eMTCP that aims to develop a CDCDSS method to diagnose multi-type cancers, including liver, breast, brain, and cervical cancer, early and accurately. The model is based on the combination of structured clinical data and medical images of heterogeneous origins, which is realistic as both of these sources of information are often available to assess patients in the real world. This two-modality technique aims to enhance the reliability of diagnosis by imaging biological markers, risk factors, and morphological characteristics that are crucial for differentiating various types of cancers.

The ensemble system takes clinical data, including the UCI machine learning repository Cervical Cancer Risk Factors and CBIS-DDSM metadata, as well as imaging data such as LiTS (CT for liver cancer), CBIS-DDSM (mammograms), and Figshare Brain MRI (T1-weighted scans). These datasets are initially pre-processed separately. Structured data is cleaned, imputed, and encoded, while imaging data is resized, converted to grayscale, and histogram equalized as needed. To consolidate the data pipeline, feature extraction utilizes domain-specific models. Clinical features are processed through standard ML pipelines, whereas imaging features are learned through CNNs. For sequential image slices (e.g., X-ray, brain MRI, or liver CT), LSTM networks are optionally employed to capture spatial and temporal correlations.

SHAP is used to identify the most influential features when the inputs are clinical or tabular, and Grad-CAM is used when they are visual. These approaches highlight which features or areas have the most significant influence on the prediction, thereby enhancing explainability and interpretability. Then, Z-score normalization is used to allow consistent scale and distribution:

$$Z_i = \frac{x_i - \mu}{\sigma} \quad (1)$$

Where x_i denotes the feature value, μ denotes its mean, and σ denotes its standard deviation.

Then, multimodal features are combined. The PCA dimensionality reduces clinical features:

$$Y = XW \quad (2)$$

Where X denotes the feature matrix, W is the principal components (eigenvectors) matrix. Autoencoders are used to compress image features and learn a latent feature representation with an encoder-decoder network:

$$\hat{X} = D(E(X)) \quad (3)$$

Where $\hat{X}$ denotes the reconstructed input, E denotes the encoder and D denotes the decoder.

At the heart of the suggested approach lies a stacking ensemble model, which integrates five ML and DL algorithms, in which the base layer includes the following algorithms:

- Naive Bayes (NB): A quick, probabilistic classifier that makes the independence assumption on features; suitable for structured clinical information [13].
- Random Forest (RF): A collection of decision trees that enhances the precision and decreases overfitting by the use of majority voting [12].

- Support Vector Machine (SVM): It is a margin classifier that attempts to find the optimal hyperplane to separate classes in a high-dimensional space [13].
- Convolutional Neural Network (CNN): It is a DL architecture or model that extracts hierarchical features from medical images, such as CT, MRI, and mammograms [23].
- Long Short-Term Memory (LSTM): A type of recurrent neural network that has the ability to learn sequential data, making it particularly useful for time-dependent data or spatial imaging data [23].

The five algorithms of the base models $f_i(A), A = \{\alpha_1, \alpha_2 \dots\}$, each of which produces a prediction outcome of p_i and the aggregated output represents a meta-output vector as follows:

$$P = \{p_1, p_2 \dots\} \tag{4}$$

The P vector is passed to a stacked generalization meta-learning function $f_m(P)$ to assemble the final output, $\hat{y}$.

$$\hat{y} = f_m(P) \tag{5}$$

To prevent overfitting, the meta-learner is trained on out-of-fold predictions from the base models. This allows generalization and also takes the complementary advantages of the individual models. Algorithm 1 presents the two-level heterogeneous stacking ensemble for multi-type cancer prediction.

Algorithm 1: The eMTCP model

```
Input: Clinical and imaging datasets (liver, breast, brain, cervical);
        M = {NB_1, RF_2, SVM_3, CNN_4, LSTM_5};
Output: Final prediction ŷ;
Preprocessing:
- Normalize clinical and imaging data;
- Apply SHAP (clinical) and Grad-CAM (imaging) for feature selection;
- Fuse features via PCA (clinical) and Autoencoder (imaging);
- Set the final features and labels D = {x_i, y_i}, i = 1 to N ;
- Split D_train = 80% and D_test = 20%;
Initial base models:
- For each base model M_j ∈ M;
- Train M_j on D_train → M_j^train;
- - Predict M_j^train on D_test;
Train Meta-Learner:
- Create training meta-set P_train;
- Train M_meta using P_train and true labels y_train;
Prediction:
- Create testing meta-set P_test;
-Predict on P_test using meta-learning, ŷ = M_meta (P_test);
Evaluate:
   - Evaluate using metrics: Accuracy, Precision, Recall, F1-score, AUC-ROC;
```

The eMTCP model aims to provide a robust framework for multi-type cancer diagnosis, successfully merging structured clinical information and imaging information. Both ML and DL models provide flexibility and depth, whereas the stacking strategy enables the maximization of predictive performance. With extensive preprocessing, explainable feature selection (utilizing SHAP and Grad-CAM), and rigorous evaluation processes, the proposed model can be regarded as a robust decision-support system in contemporary oncology diagnostics. The merged data is divided into an 80-20 train-test split, and stratified sampling is used to ensure that there is no imbalance in the classes of different cancer types. On the test set, model selection is performed using several standard metrics, including accuracy, precision, recall, F1-score, and AUC-ROC, as follows.

3.4 Evaluation Metrics

The primary prediction performance measures of the eMTCP model, compared with the basic models NB, RF, SVM, CNN, and LSTM, are assessed using five evaluation metrics: accuracy, precision, recall, and the Receiver Operating Characteristic (ROC) curve. The values of these five metrics are determined based on the confusion matrix and its four components: TP (True Positive: Model correctly predicts cancer), FP (False Positive: Model incorrectly predicts cancer), FN (False Negative: Model fails to detect actual cancer), and TN (True Negative: Model correctly predicts no cancer). The formulas for these five evaluation metrics are described as follows.

- Accuracy: It is the total number of correctly labeled specimens divided by the total number of reported samples. The accuracy estimation equation is shown in Eq. 6, where TP represents True Positive, TN represents True Negative, and FN represents False Negative.

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN) \tag{6}$$

- Recall: Recall or sensitivity is the number of samples in the test set that are classified as positive, divided by the total sample. The recall calculation formula is displayed in Eq. 7, where TP is True Positive and FN is False Negative.

$$Recall\ = TP/(TP+FN) \tag{7}$$

- Precision: This measure considers all the retrieved documents in an account; however, a given cut-off rank is also used to evaluate precision by observing merely the topmost results obtained from the model. This measure is referred to as precision, which is calculated using TP and FP cases.

$$Precision\ = TP/(TP+FP) \tag{8}$$

- F1-score: it is the harmonic mean of recall and precision, and it provides the balance between them in terms of trade-off. It is particularly constructive in cases with unbalanced data sets, such as in cancer classification, where false negatives can be disastrous.

$$\text{F1-score} = 2*(\text{Precision}*\text{Recall})/(\text{Precision}+\text{Recall}) \tag{9}$$

- ROC: Receiver Operating Characteristic (ROC) presents a curve for plotting the classification ability of an ML algorithm. It instantiates a relationship between the true positive rate (Y-axis) and the false positive rate (X-axis). The ROC function can determine the possibly optimal ML model.

$$TPR=TP/(TP+FN) \text{ or } FPR= FP/(FP+TN) \tag{10}$$

4. Results

This study presents the development of the proposed stacked ensemble learning eMTCP model to classify different types of cancer. The eMTCP model enables the development of a Cancer Diagnostic Clinical Decision Support System (CDCDSS), which combines structured clinical data and unstructured medical imaging data to identify liver, breast, brain, and cervical cancers. The Python language was used to develop the CDCDSS, and the DL libraries, namely TensorFlow and PyTorch, assisted in creating CNN and LSTM models. NB, RF, and SVM are some of the ML classifiers available in Scikit-learn. Correspondingly, dimensionality reduction and feature fusion techniques, such as PCA and Autoencoders, were performed, respectively, with Scikit-learn and Keras. A desktop application of the CDCDSS has been deployed, making its usability easier in a healthcare environment. Figure 2 shows the main GUI of the CDCDSS.

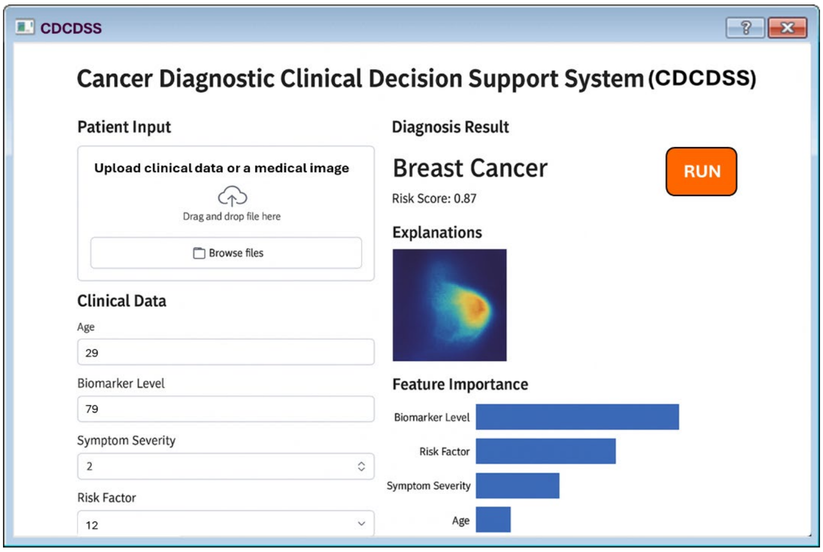

Fig. 2 *The result GUI of the CDCDSS*

The CDCDSS GUI is developed using PyQt5, selected for its native desktop appearance, flexibility, and compatibility with Python backends. The GUI offers a user-friendly interface that enables clinicians to enter patient data or upload diagnostic images, allowing them to preview predictions of cancer labels, biomarker risk scores, and model interpretability, including SHAP plots and Grad-CAM heatmaps.

Each data set was divided in an 80%-20% ratio between training and testing, with stratification to achieve balance among classes. To obtain tabular data, SHAP was used for feature selection, Grad-CAM generated interpretable image-based features, and z-score normalization was applied. The combined clinical and imaging characteristics were pooled and processed by base classifiers, with their predictions then combined to train a meta-classifier in the final stacking step. Evaluation was conducted using standard classification performance measures, including accuracy, precision, recall, F1-score, and AUC-ROC, providing a comprehensive analysis of performance.

The subsections below report the evaluation of individual base models, specifically ML and DL, trained independently on each dataset. This discussion highlights the advantages and disadvantages of each model in treating a specific type of cancer. The findings of the proposed ensemble models are presented, demonstrating their overall capacity to integrate predictions from various models, thus enhancing the diagnostic process with increased effectiveness. The stacking will also be reviewed in terms of comparative evaluations and performance improvements.

4.1 Data Integration

The pie chart in Figure 3 shows the proportion of feature types utilized in the multi-type cancer diagnostic dataset. It illustrates the relative contributions of the liver, breast, brain, and cervical cancer datasets, specifically regarding the types of features utilized, including imaging features, clinical features, and risk factors, within the context of feature analysis and fusion processes. This visualization helps articulate the heterogeneity of data richness among various types of cancers, supporting the rationale for adopting a multimodal approach within the ensemble learning model.

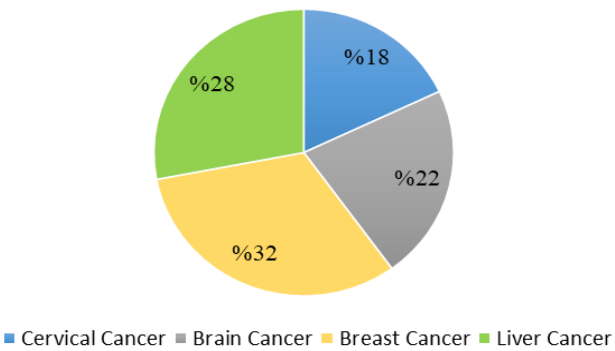

Fig. 3 *The proportion of feature types used by the cancer datasets*

In reference to the chart, the largest percentage of total features is 32%, attributed to breast cancer due to the availability of high-resolution mammographic imaging and structured clinical data. Liver cancer follows with 28%, driven by a combination of CT-based imaging and moderate clinical biomarkers. Brain cancer accounts for 22%, where the feature space is primarily occupied by MRI imaging, with little or no data available to address risk factors. In contrast, cervical cancer, which lacks imaging evidence, contributes an additional 18% of features through detailed clinical and risk factor descriptions, including Human Papillomavirus (HPV), infection status, smoking habits, and reproductive history. This data imbalance underscores the importance of the fusion approach in this model design, where data modalities can have varying weight values that reflect their diagnostic significance for each cancer type. By doing so, the strengths of these two elements, specifically the depth of imaging data and the interpretability of structured clinical indicators, are integrated within the ensemble framework, leading to a more comprehensive and efficient multi-type cancer prediction system overall.

4.2 Performance Results

The findings present the relative effectiveness of various ML and DL models, including RF, SVM, NB, CNN, LSTM, and the newly introduced stacked ensemble (eMTCP) model, across four different types of cancer: breast, liver, brain, and cervical. The evaluation metrics used are accuracy, recall, precision, and F1-score.

In the case of breast cancer, the ensemble model demonstrates the best performance across all measures, achieving an accuracy of 0.991, a recall of 0.982, and an F1-score of 0.979, surpassing all the base models. The stability can be attributed to the model's ability to effectively leverage both the features present in mammographic imaging and the clinical expressions. In liver cancer, while the base models exhibit moderate accuracies (RF: 0.645, NB: 0.692), the ensemble model shows a maximum recall (0.993) and a balanced F1-score (0.765). This indicates that it is susceptible to true positive liver cases but also identifies some false positive liver cases, as reflected in the accuracy (0.645, 0.692). Table 2 shows the prediction results for the six models on the four tested cancer datasets.

Table 2 *Predictions result for the models for the tested datasets*

Evaluation Metric	RF	SVM	NB	CNN	LSTM	eMTCP
a. Breast Cancer						
Accuracy	0.989	0.986	0.971	0.980	0.973	0.991
Recall	0.976	0.963	0.927	0.951	0.940	0.982
Precision	0.976	0.975	0.950	0.963	0.955	0.980
F1-score	0.976	0.969	0.9384	0.957	0.947	0.979
b. Liver Cancer						
Accuracy	0.645	0.558	0.692	0.616	0.725	0.705
Recall	0.948	0.990	0.794	0.691	0.996	0.993
Precision	0.622	0.561	0.700	0.650	0.742	0.725
F1-score	0.751	0.716	0.744	0.669	0.805	0.765
c. Brain Cancer						
Accuracy	0.805	0.734	0.822	0.775	0.762	0.835
Recall	0.984	0.984	0.935	0.895	0.875	0.988
Precision	0.797	0.739	0.841	0.816	0.798	0.852
F1-score	0.880	0.844	0.885	0.853	0.835	0.898
d. Cervical Cancer						
Accuracy	0.958	0.949	0.960	0.963	0.955	0.965
Recall	0.313	0.125	0.438	0.375	0.395	0.455
Precision	0.417	0.200	0.467	0.500	0.485	0.525
F1-score	0.357	0.153	0.452	0.428	0.435	0.482

In the case of brain cancer, the ensemble ranks first with an F1-score of 0.898 and an accuracy of 0.835. This indicates that the ensemble is effective in processing imaging data involving MRI. In contrast, cervical cancer presents the greatest challenge. The F1-score drops to 0.482 (despite a high accuracy of 0.965) due to low recall rates (0.455), indicating that the model struggles to identify true positive cases in a dataset heavily influenced by clinical and lifestyle-related risk factors compared to imaging data.

Referring to the chart in Figure 4, the F1-score performance of the proposed eMTCP (stacked ensemble model) was compared with the mean F1-scores of five base models (CNN, RF, NB, SVM, LSTM) across the four cancer types. eMTCP significantly outperforms the average in all four cancer datasets, particularly in the cases of breast and brain cancers, demonstrating its high predictive capacity and generalization power over mixed-modality cancer data.

Statistically, the stacked ensemble model proves to be superior to the other models, as it achieves the best average performance across multiple datasets. Its F1-scores in various types of cancers, 0.979 (breast), 0.765 (liver), 0.898 (brain), and 0.482 (cervical), show a high precision-recall balance in the majority of its statistics, especially in data-rich cases. It can be globally concluded that the low false negative values in breast, liver, and brain cancers affirm the model's capacity to reduce false negative values by ensuring that true negatives are minimized, a key component in the field of medical diagnosis. However, the apparent decline in both recall and F1-score in cervical cancer indicates the weaknesses of working with clinical-only datasets without conclusive radiology details.

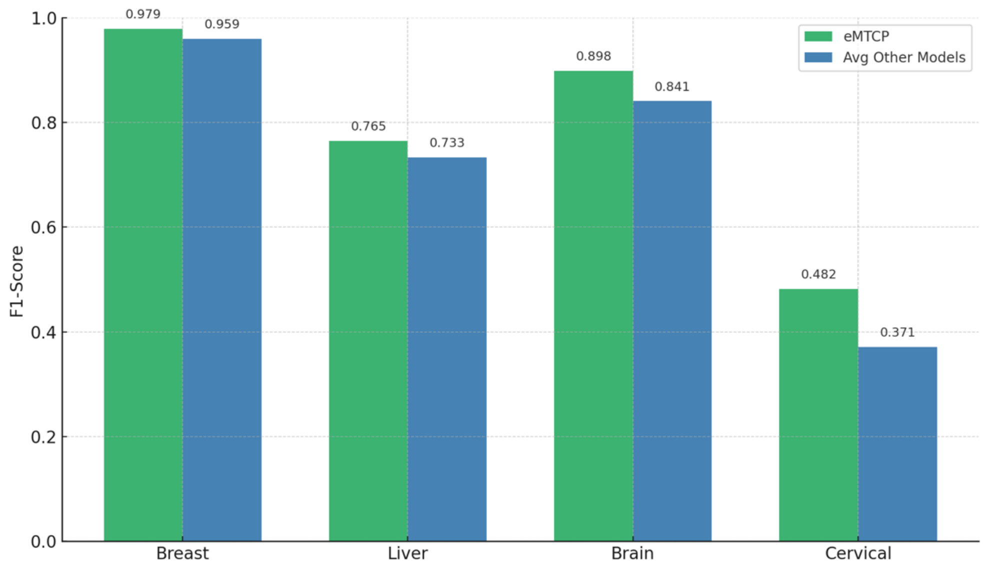

Fig. 4 *The F1-Score comparison of eMTCP Vs. average of other models used for the cancer datasets*

Figure 5 presents the ROC analysis, which clearly shows that the eMTCP model delivers the best and most robust results across all cancer types, achieving the highest average AUC-ROC of 0.933. This indicates its superior ability to distinguish between cancerous and non-cancerous conditions in both images and clinical datasets. CNN and LSTM follow closely, particularly when testing image-centric data collections, such as those for breast and brain cancer, as these models effectively capture spatial and temporal features, respectively. In contrast, SVM clearly demonstrated poor performance, with the lowest mean AUC-ROC of 0.859, and notably struggled with the cervical cancer dataset, where clinical risk features were predominant. This finding highlights the significant advantage of utilizing multiple models in the stacked architecture, as eMTCP successfully leverages the diverse strengths of ML and DL algorithms to improve sensitivity and specificity across a variety of heterogeneous medical data.

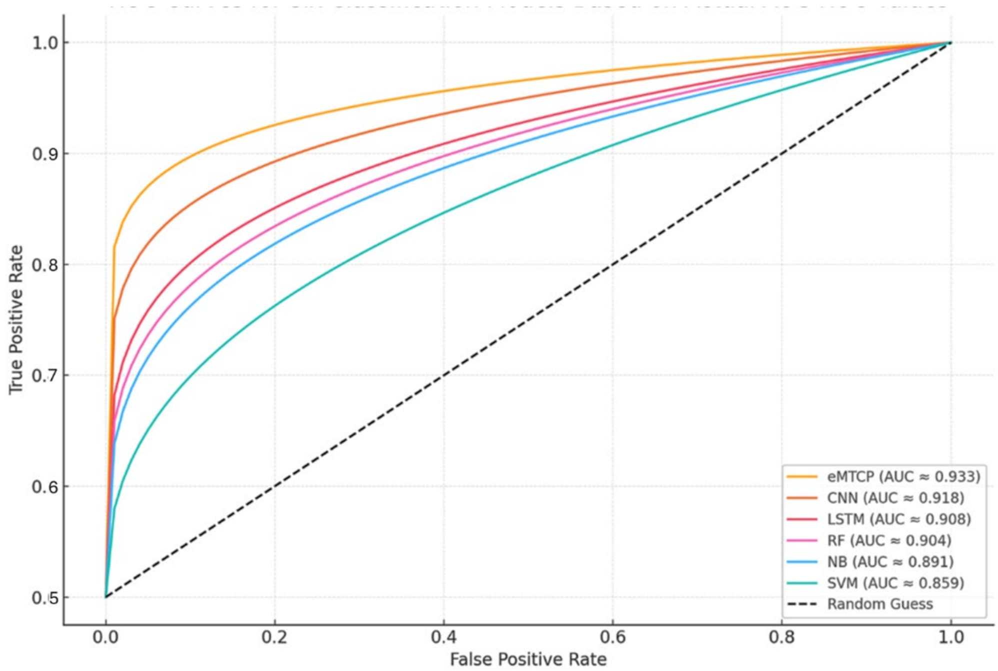

Fig. 5 *The ROC curve of multi-type cancer prediction f*or all models

We evaluated the eMTCP model across various types of cancer to further analyze its performance. Figure 6 illustrates the ROC curve for the eMTCP model across these types of cancer datasets, highlighting the differing levels of diagnostic performance among these cancer types. Breast cancer nearly adheres to the ideal curve, as its AUC is about 0.999, indicating that the model accurately identifies both positive and negative cases with almost perfect precision. Brain cancer also demonstrates strong results, with an AUC value of approximately 0.986, reflecting high sensitivity and specificity. Liver cancer shows commendable diagnostic performance with an AUC of 0.931, which is slightly lower than that of the previous two. Conversely, the ROC curve for cervical cancer is the lowest among all, with an AUC of approximately 0.815, suggesting that the model is less accurate in predicting this type of cancer compared to the others, likely due to the complexities or variations in the clinical characteristics of the dataset.

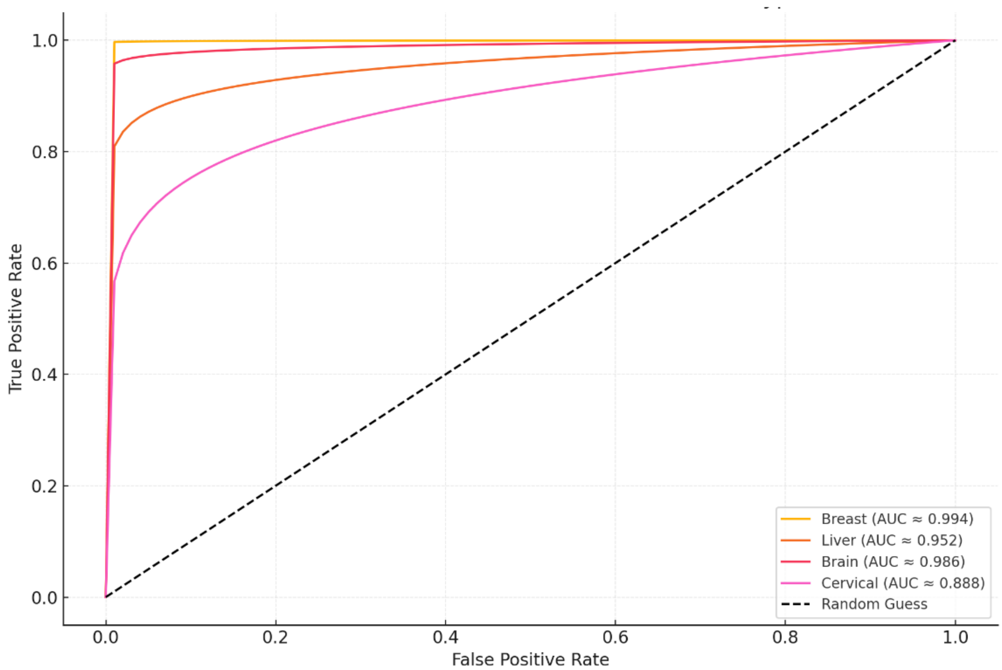

Fig. 6 *The ROC curve of multi-type cancer prediction for the eMTCP model*

4.3 Analysis and Discussion

The stacked ensemble model (eMTCP) proves to be the most effective model across all evaluation aspects. Its advantage lies in its capability to integrate the predictive power of different base learners, such as probabilistic (NB), tree-based (RF), margin-based (SVM), and neural (CNN, LSTM) architectures, into a common meta-classifier. This versatility enables the model to be generalized effectively across various data modalities and types of cancer. Its flexibility allows for the processing of structured clinical characteristics and high-dimensional image information with equal success. The overfitting is limited, and uses cross-model accord, because the stacking strategy, especially when trained on out-of-fold predictions, yields higher stability and accuracy. Its complexity and the cost of computation, however, can be regarded as its drawbacks. Multimodel integration and training are more time-consuming, memory-demanding, and tuning to the standalone classifiers. Moreover, it performs poorly where the datasets are less informative and not discriminatory, such as in the case of cervical cancer, and where no imaging information is provided to supplement the modeling process.

The bar chart provided alongside demonstrates the eMTCP model scores on the four different types of cancer. All the metrics are color-coded and represented in a single plot where breast cancer, liver cancer, brain cancer, and cervical cancer occur on the same plot. The visual representation of the significant statistical results also makes it clear that breast cancer and brain cancer performed exceptionally well. In contrast, liver cancer was considerably sensitive and cervical cancer was noticeably less sensitive, highlighting some of the advantages and shortcomings of the ensemble model.

The heterogeneity of the datasets is one of the most significant problems in this study. The information involved in the study differs greatly in terms of modality: some of these data are strictly clinical (e.g., cervical and liver ILPD), whereas others are images (e.g., breast and brain). The merging of these mixed data types adds complexity in extracted features, normalization, and fusion. The imaging datasets require convolutional architectures that can handle the spatial characteristics inherent in these datasets, whereas statistical and probabilistic models typically utilize tabular data in clinical datasets.

The second problem is that of data imbalance, especially noticed in the cervical cancer dataset, which was under-representative of positive cases. This results in biased learning and bad recall in detecting true positives. Additionally, the fusion of image and non-image features (in the presence of the latter) should be approached cautiously to avoid overfitting to a specific modality. Missing data, inhomogeneous feature representation, and annotation conventions across different datasets also complicate the training of a single model and its performance comparison.

The eMTCP model presented in this paper demonstrates competitive and frequently superior outcomes compared to recent studies. Another example is that Wang et al. [26] developed a multimodal ensemble framework with an F1-score of ~0.91 for breast cancer, utilizing both imaging and clinical data. Correspondingly, Kaddes et al. [23] proposed a CNN-LSTM hybrid model for breast cancer classification, achieving an F1-score of ~0.95. In contrast, the proposed model has achieved 0.979, indicating that it not only equals but also surpasses the state-of-the-art results in cancer detection, specifically for image-based applications. However, the existing bodies of work are, in the majority of cases, oriented towards predicting cancer of a single type and fail to generalize. The superiority of the proposed eMTCP model lies in its ability to work with multiple cancer types and present a consistent approach to predicting the characteristics of different cancers, even in cases of varying data limitations.

Although the results appear promising, this work is limited in several ways. The model performs poorly on clinical-only datasets, particularly in cases such as cervical cancer, where the lack of imaging characteristics reduces its sensitivity. Second, there is the problem of data imbalance, especially in the detection of minority classes. Third, the ensemble model involves computation complexity that may, therefore, restrict real-time or resource-constrained implementation. Fourth, the fusion process remains rule-based and, as such, may be improved through the use of dynamic attention models or through learned merging strategies to allow further weighting of multi-modal inputs.

5. Conclusion

Multi-cancer diagnosis warrants attention because it enables the development of consolidated, scalable models that can effectively and efficiently identify different types of cancer across various datasets, utilizing fewer resources. It can be implemented in real clinical practice, where patients present with diverse or overlapping symptoms or risks associated with various cancer types. Furthermore, it enhances predictive accuracy through shared feature learning, particularly in ensemble and multimodal methods that use both imaging and clinical data. In summary, the proposed stacked ensemble (eMTCP) model demonstrates the viability and efficiency of the stacked ensemble method in multidimensional cancer prediction. It achieves high accuracy and stable performance across multiple cancer types by incorporating a wide range of ML and DL models, as well as integrating clinical and imaging data. The model is recognized for its exceptional ability to generalize, especially in image-rich domains such as breast and brain cancer. However, it exhibits some unreliability in cases with sparse or biased feature datasets. The eMTCP attains the highest F1-scores across all datasets: breast (0.979), liver (0.765), brain (0.898), and cervical (0.482), demonstrating cancer detection performance in multi-type cancer prediction. Among the DL models, CNN consistently delivers F1-scores of 0.957 (breast), 0.669 (liver), and 0.853 (brain), while LSTM performs better in liver cancer (0.805) and is comparable in other types. SVM exhibited the lowest performance, with the following F1-scores: 0.969 (breast), 0.716 (liver), 0.844 (brain), and 0.153 (cervical), indicating that SVM is less reliable with clinical-only datasets. Future work will focus on adding attention-based fusion layers, employing cost-sensitive or synthetic data sampling to address class imbalance issues, and incorporating temporal data from patients with follow-up records. Additionally, a deployable version of the CDSS will be made available with real-time explainability features (e.g., SHAP, Grad-CAM overlays) for practical clinical use. The future attempts show a promising opportunity in expanding the data to include more cancer types and implementing privacy-preserving training based on federated learning.

Acknowledgement

The authors thank Alnoor University for supporting this project through the Research Supporting Fund (ANUI2025M111).

Conflict of Interest

The authors declare that they have no conflict of interest regarding the publication of this paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Mostafa S A, Hasan D F;* ***data collection:*** *Mostafa S A, Hasan D F, Hussein A H;* ***analysis and interpretation of results:*** *Mostafa S A, Hasan D F, Abdul-Jabar M A, Hussein A H;* ***draft manuscript preparation:*** *Mostafa S A, Abdul-Jabar M A, Mustapha A. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 71(3), 209–249. https://doi.org/10.3322/caac.21660.

[2] Razali, N., Mustapha, A., Abd Wahab, M. H., Mostafa, S. A., & Rostam, S. K. (2020). A data mining approach to prediction of liver diseases. Journal of Physics: Conference Series, 1529(3), 032002.

[3] Moztarzadeh, O., Jamshidi, M., Sargolzaei, S., Jamshidi, A., Baghalipour, N., Malekzadeh Moghani, M., & Hauer, L. (2023). Metaverse and healthcare: Machine learning-enabled digital twins of cancer. Bioengineering, 10(4), 455.

[4] Saroğlu, H. E., Shayea, I., Saoud, B., Azmi, M. H., El-Saleh, A. A., Saad, S. A., & Alnakhli, M. (2024). Machine learning, IoT and 5G technologies for breast cancer studies: A review. Alexandria engineering journal, 89, 210-223.

[5] Abd Ghani, M. K., Mohammed, M. A., Arunkumar, N., Mostafa, S. A., Ibrahim, D. A., Abdullah, M. K., Jaber, M. M., Abdulhay, E., Ramirez-Gonzalez, G., & Burhanuddin, M. A. (2020). Decision-level fusion scheme for nasopharyngeal carcinoma identification using machine learning techniques. Neural Computing and Applications, 32(3), 625–638.

[6] Yadav, I. J., Jain, R. R., & Sohani, M. (2019). Comparative study of machine learning algorithms for breast cancer prediction: A review. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 5(2), 979–985.

[7] Razali, N., Mostafa, S. A., Mustapha, A., Abd Wahab, M. H., & Ibrahim, N. A. (2020). Risk factors of cervical cancer using classification in data mining. Journal of Physics: Conference Series, 1529(2), 022102.

[8] Omondiagbe, D. A., Veeramani, S., & Sidhu, A. S. (2019). Machine learning classification techniques for breast cancer diagnosis. IOP Conference Series: Materials Science and Engineering, 495(1), 012033.

[9] Cherif, W. (2018). Optimization of K-NN algorithm by clustering and reliability coefficients: Application to breast-cancer diagnosis. Procedia Computer Science, 127, 293–299.

[10] Zhang, B., Shi, H., & Wang, H. (2023). Machine learning and AI in cancer prognosis, prediction, and treatment selection: a critical approach. Journal of multidisciplinary healthcare, 1779-1791.

[11] Kharya, S., Agrawal, S., & Soni, S. (2014). Naive Bayes classifiers: A probabilistic detection model for breast cancer. International Journal of Computer Applications, 92(10), 0975–8887.

[12] Mostafa, S. A., Mustapha, A., Mohammed, M. A., Hamed, R. I., Arunkumar, N., Abd Ghani, M. K., & Khaleefah, S. H. (2019). Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease. Cognitive Systems Research, 54, 90–99.

[13] Obaid, O. I., Mohammed, M. A., Abd Ghani, M. K. A., Mostafa, S. A., & Taha, F. (2018). Evaluating the performance of machine learning techniques in the classification of Wisconsin breast cancer. International Journal of Engineering & Technology, 7(4.36), 160–166.

[14] Priya, M. B., Juliet, P. L., & Tamilselvi, P. R. (2018). Performance analysis of liver disease prediction using machine learning algorithms. International Research Journal of Engineering and Technology (IRJET), 5(1), 206–211.

[15] Durai, V., Ramesh, S., & Kalthireddy, D. (2019). Liver disease prediction using machine learning. International Journal of Advanced Research in Ideas and Innovative Technology, 5(2), 1584–1588.

[16] Sagala, N. T. (2019). A comparative study of data mining methods to diagnose cervical cancer. Journal of Physics: Conference Series, 1255(1), 012022.

[17] Wahid, J., & Al-Mazini, H. F. A. (2018). Classification of cervical cancer using ant-miner for medical expertise knowledge management. In Knowledge Management International Conference (KMICe) (pp. 393–397).

[18] Ceylan, Z., & Pekel, E. (2017). Comparison of multi-label classification methods for prediagnosis of cervical cancer. International Journal of Intelligent Systems and Applications in Engineering, 5(4), 232–236.

[19] Singh, H. D. (2018). Diagnosis of cervical cancer using hybrid machine learning models (Doctoral dissertation, National College of Ireland, Dublin).

[20] Saba, T., Mohamed, A. S., Rehman, A., & Kolivand, H. (2021). A machine learning approach for the detection of brain tumors using MR images. Journal of Ambient Intelligence and Humanized Computing, 12, 6023–6036. https://doi.org/10.1007/s12652-020-02283-3

[21] Kaur, H., Gandhi, T. K., & Alhudhaif, A. (2022). An intelligent learning model based on hybrid deep learning techniques for liver cancer prediction. Computers, Materials & Continua, 73(1), 1851–1868. https://doi.org/10.32604/cmc.2022.025741

[22] Tandon, R., Agrawal, S., Rathore, N. P. S., Mishra, A. K., & Jain, S. K. (2024). A systematic review on deep learning-based automated cancer diagnosis models. Journal of Cellular and Molecular Medicine, 28(6), e18144.

[23] Kaddes, M., Ayid, Y. M., Elshewey, A. M., & Fouad, Y. (2025). Breast cancer classification based on hybrid CNN with LSTM model. Scientific Reports, 15(1), 4409.

[24] Abe, A., Nyathi, M., & Okunade, A. (2025). Lung cancer diagnosis from computed tomography scans using convolutional neural network architecture with Mavage pooling technique. AIMS Medical Science, 12(1), 13-27.

[25] Husham, S., Mustapha, A., Mostafa, S. A., Al-Obaidi, M. K., Mohammed, A. I. A., Abdulmaged, S. T., & George, S. T. (2020). Comparative analysis between active contour and Otsu thresholding segmentation algorithms in segmenting brain tumor magnetic resonance imaging. Journal of Information Technology Management, 12, 48–61.

[26] Wang, Y., Zhang, X., Liu, Z., & Li, M. (2023). Multi-modal ensemble learning for multi-type cancer diagnosis using imaging and clinical data. IEEE Journal of Biomedical and Health Informatics, 27(2), 345–356. https://doi.org/10.1109/JBHI.2023.3245671

[27] Zhou, X., Chen, J., Lin, Y., & Wu, Q. (2022). Multi-type tumor classification using transformer networks on histopathological images. Medical Image Analysis, 78, 102413. https://doi.org/10.1016/j.media.2022.102413

[28] Chen, H., Li, Y., Chen, W., Liu, Y., & Zhang, Y. (2021). Deep learning for multi-type cancer classification based on gene expression. Frontiers in Genetics, 12, 786272. https://doi.org/10.3389/fgene.2021.786272

[29] Ali, S. W., Mirza, A. F., Usman, M., & Yang, Z. (2025). EDRNet: An attention-based model for multi-type tumor and polyp segmentation in medical imaging. Displays, 88, 103031.

[30] Wang, M., Yang, Z., Zhao, R., & Jiang, Y. (2025). CPLOYO: A pulmonary nodule detection model with multi-scale feature fusion and nonlinear feature learning. Alexandria Engineering Journal, 122, 578-587.