https://doi.org/10.48048/tis.2023.6299

A Label-Free Marker Based Breast Cancer Detection using Hybrid Deep Learning Models and Raman Spectroscopy

Renjith Vijayakumar Selvarani* and Paul Subha Hency Jose

Department of Biomedical Engineering, Karunya Institute of Technology and Sciences, Tamil Nadu 641114, India

(*Corresponding author's e-mail: notify_renjithvs@yahoo.com)

Received: 29 October 2022, Revised: 8 December 2022, Accepted: 15 January 2023, Published: 22 January 2023

Abstract

Breast Cancer (BC) is a serious menace to women's health around the world. Early BC identification has been critically important for diagnosing protocol. Several classification methods for breast cancer were examined recently with various techniques, and Raman spectroscopy (RS) has become an effective approach for the identification of responsible metabolites. Moreover, the rapid and accurate classification of BC using RS necessitates active engagement in processing and analyzing Raman spectral data. This work aims to develop an efficient Hybrid Deep Learning (HDL) neural network model to differentiate breast cancer blood plasma from control samples and the spectral features obtained are used as spectral cancer markers for the detection of breast cancer. To find the optimum performing HDL model, several other HDL models were implemented to perform the binary classification of the Raman spectral signal. A total of 62199 Raman spectra generated from 26 blood plasma samples are evaluated in this study. Mainly 6 HDL methods, 1D-CNN-GRU, CNN-BiLSTM-AT, 1D-CNN-LSTM, GRU-LSTM, RNN-LSTM, and OGRU-LSTM are modeled to evaluate the performance of hybrid models to identify 2 classes of Raman spectral data. Comparative classification results show that the stacked 1D-CNN-GRU model outperforms well for breast cancer detection using the Raman spectral dataset than other prominent HDL architectures. The stacked 1D-CNN-GRUclassifier model achieved the highest classification accuracy (98.90 %), Cohen-kappa score (0.941), F1-score (0.969), and the lowermost number of test loss as 0.102776 and MSE (0.0230) indicating that the model outperforms other HDL classifiers.

Keywords: Breast cancer detection, Hybrid deep learning, Spectral data augmentation, 1Dimensional-CNN-GRU

Introduction

Breast Cancer causes the foremost risk to women's health, accounting for 24 % of all types of cancer in women, thus BC is the leading cause of death of 15 % of all types of cancer [1]. Detection and accurate classification of BC at the early stage can reduce the complexity of medical interventions and the mortality rate. X-ray mammography is the existing gold standard for screening breast mass, and it is widely employed for the diagnosis of BC [2]. Moreover, mammography functions poorly in dense breasts and can't differentiate between malignant and benign [3], and in the case of ultrasound and magnetic resonance imaging (MRI) techniques, the limitations are the same. As a result, a needle or surgical excision biopsy is still required for further diagnosis, extending the diagnostic procedure by weeks and increasing the patient stress and costs of medical aids.

Raman spectroscopy is an extensively used method for the identification of organic and inorganic chemicals. Major enhancements in lasers, holographic optical components, spectrometers, and detectors have high-pointed the significance of Raman spectroscopy as an effective and productive device for a variety of applications including bio-science, medical diagnostics, planetary investigations, and monitoring of the in-situ process. In clinical research, Raman spectroscopy is a noninvasive optical imaging and spectroscopic technology that can enhance the diagnosis of cancer with less time, higher sensitivity, and better specificity. It is a fingerprint spectroscopic technique that analyses inter-molecular functional groups and their related molecular structures by measuring vibrations and rotational energy. Due to its non-destructive aspect, Raman spectroscopy can enable quick molecular analysis of tissues *in vitro* or *in vivo* for laboratory purposes or biopsy [4]. The molecular contents, morphologies of BC, normal cells, and

tissues can be detected using Raman spectroscopy [5-7]. The capability to identify benign from malignant tissues using Raman spectra has been proven in large samples of blood [8]. However, due to the intricacy of intramolecular components and chemical structures, Raman spectra from bio-samples carry a lot of biological information. As a result, the biomolecules' active functional groups relating to Raman spectra with diverse molecular and chemical properties yield different patterns of the spectrum, which can be utilized as "fingerprints" of markers significant to certain BC. Moreover, variations within those "fingerprints" can convey information about the disease, which is critical for disease diagnosis and tracking [9]. Raman spectroscopy has demonstrated its significance in the domains of bio-macromolecules identification, detection of pathogenic micro-organisms, tumor diagnosis, and other similar fields [10-12] due to its specific technological benefits of non-destructive testing methodology, high sensitivity, and speed. The computer-aided deep learning (DL) based data discovery algorithms are employed to automate the spectral analytical system thus allowing healthcare professionals to participate in the analysis of BC from the spectrogram without any manual intervention. The deep learning approaches along with RS can analyze the variations in chemical compositions that exist among different samples based on the morphology. It can identify variations in molecular structure and composition in the constituents of tissue and cell such as proteins, carbohydrates, lipids, and nucleic acids that occur during tumor growth. The metabolic products of tumors are carried by the blood and can be tracked during circulation. Thereby analyzing the structural differences and molecular composition of the biomarkers circulated within blood samples effectively by employing the deep learning classifiers at the molecular level. The ability of RS to assist in the diagnosis of early-stage BC [14], is potent since these biochemical variations manifest [15], before the appearance of clinical signs frequently recognized by imaging protocols. DL techniques, which are an ensemble form of machine learning, are frequently beneficial due to their self-adaptive structure, which can analyze spectral data with less computational processing. Though, considering the limits of current cancer detection methods using Raman spectroscopy and conventional machine learning approaches, particularly early-stage cancer detection, there is an immediate necessity to develop novel methods for rapidly diagnosing cancer at a very early stage from the spectrogram generated by Raman spectroscopy with correlating features.

Label-free RS along with DL models are extensively used for the early cancer screening and staging of cancer studies in recent years. However the deep learning architecture works well in the classification of time series spectral patterns, they are not significant and effective in terms of reliability, robustness, steadiness in performance, and classification accuracy with a large number of higher dimensional spectral patterns with a higher number of spectral attributes associated with the spectral signature when comparing with ground truth values obtained from the clinical diagnosis, also the Raman spectral signals are weak and get easily affected by the fluorescence backgrounds since the cross-section of scattering of some molecules is quite small, resulting in a negligible differential in the spectra of normal and cancerous/disease blood samples [13]. To identify these samples, complex statistical analysis approaches are required to extract useful information from the extensive Raman spectrums. The novelty of the proposed study is the development of hybrid deep learning models (HDL) by combining the most essential aspect of different learning models. The hybrid deep learning (HDL) approach involves the use of various types of deep neural networks, such as CNN, GRU, BiLSTM, RNN, and OGRU, in conjunction with probabilistic methods to account for uncertainty in the model. These deep learning algorithms have demonstrated strong performance and have been widely applied to various types of time series data. However, they do not explicitly consider uncertainty, unlike HDL models which incorporate both deep learning and probabilistic techniques to take advantage of the strengths of both approaches. This ensemble model aid in the analysis by extracting the deep characteristic features of datapoints, particularly the breast cancer characteristics of Raman spectral fingerprints. HDL performs better in the detection of BC at early stages in asymptomatic candidates with high accuracy in real-time with a minimal quantity of blood samples pertaining to the Raman spectral patterns associated and clinical observations made by utilizing the strength of ensemble deep learning models. This would potentially lead to the identification of BC at an early stage, thus allowing timely diagnosis and treatment to save lives. Compared to other DL methods in terms of performance, HDL neural network architecture may scale exponentially to account for the increasing volume and dimensionality of spectral data. This makes HDL mostly useful for resolving complex computational difficulties such as large-scale time series spectral signal classification.

In this paper, hybrid deep learning architecture-based systems are modeled and proposed, and their performances on the spectral data are analyzed to provide a robust model for improved accuracy for BC prediction and classification by employing Raman spectral data. Several deep learning hybrid models are discussed in this explorative research work to find the effectiveness of increased prediction accuracy and learning efficiency of the deep learning hybrid models. Quantitative, comparative, qualitative, and

complexity measurements are all used in the analysis of the algorithms proposed. Finally, this work proposes an efficient Hybrid Deep Learning (HDL) neural network model to differentiate breast cancer blood plasma from control samples and the spectral features obtained can be used as spectral cancer markers for the detection of breast cancer. To find the optimum performing HDL model, several other HDL models were created to perform the binary classification of the spectral signal.

The spectral characteristics obtained from blood plasma samples from patients diagnosed with BC are analyzed using Raman spectroscopy along with the 1D-CNN-GRU (1-dimensional convolutional neural network-gated recurrent unit), CNN-BiLSTM (convolutional neural network-Bidirectional long short term memory), 1D-CNN-LSTM (1-dimensional convolutional neural network- long short term memory), GRU-LSTM (gated recurrent unit- long short term memory), RNN-LSTM (recurrent neural network- long short term memory) and OGRU-LSTM (optimized gated recurrent unit- long short term memory**)** classification models and are evaluated the performance of hybrid models to identify 2 classes of Raman spectral data and conclude which hybrid model performs better on the binary classification of Raman spectrums as cancerous or healthy.

The paper is strongly motivated for early BC prediction as:

1) To generate a balanced spectral dataset, pre-processing is performed;

2) This work addresses the data scarcity of spectral data associated with the Raman spectrogram of breast cancer blood plasma samples by using deep learning-based data augmentation;

3) Development and implementation of different hybrid deep learning models to find models' optimum performance with improved accuracy on Raman spectral data;

4) Comparison and performance evaluation of different hybrid deep learning models and propose the optimum HDL model;

5) To effectively differentiate characteristic features of breast cancerous and controlled samples and identify the spectral features responsible for breast cancer and control.

The remaining part of this paper is organized as follows. Section 2 reviews several existing methods. Section 3 particulars the sample preparation, acquisition of Raman spectral signal, preprocessing of the spectra, and finally data augmentation pertaining to this research under materials and methods. Section 4 describes the methodology and implementation, where it starts with the scope of using hybrid deep learning architectures for time series data pertaining to this research work, afterward describing 6 hybrid deep learning architectures such as 1D-CNN-GRU, RNN-LSTM, CNN-BiLSTM, 1D-CNN-LSTM, GRU-LSTM, and OGRU-LSTM. Section 5 is the results and discussion pertaining to this research, where the tested results of the classification methods have been discussed, and the results of each hybrid deep learning model are compared to conclude the optimum HDL method pertaining to this research work with maximum efficiency on Raman spectral data. The conclusion is enumerated in Section 6.

Related works

Raman spectroscopy is a spectroscopic technique that detects the vibrational modes of specimens physically. This characteristic information regarding the vibrational state is used to create fingerprints of spectral samples, which can then be utilized in a variety of applications to assess the biochemical constituents of the sample without affecting it or to detect a specimen characteristic, such as a patient's disease conditions. Moreover, the Raman signal is modest, and the Raman data is non-specific because of the label-free nature of RS. As a result, analyzing Raman spectrums are difficult, and chemometric methods based on ML-based are required. Deep learning (DL) has achieved considerable milestones in data science as a subset of ML techniques for the study of photonic data and Raman spectra in general. Recent advances in DL methods for Raman spectroscopy, as well as current difficulties in their use in the implementation, will be covered in this study. Deep learning is a subfield of artificial intelligence (AI) known as transfer learning. Feedforward neural networks (FNNs) are frequently used as the base of DL architecture, which is a type of artificial-neural network (ANN) that have an output, a hidden, and an input layer. The input-layer transmits data into the network, the next neurons in the hidden-layers process the data based on their weights, and finally, the output layer returns the processed spectral data. Backpropagation and gradient-based optimization algorithms are commonly used to update the network's weights and biases. This architecture enables a deep learning network to express functions of complexity by the addition of added units and layers, as long as enough no. of labeled training samples are provided. Various deep learning-based algorithms, such as CNNs, RNNs, ResNets, GANs, autoencoders, and others, have recently been devised and implemented based on this basic architecture [16].

When comes to the application of deep learning using Raman spectral signal associated with cancer-specific markers, not many studies were done with either HDL or deep learning to distinguish spectral

patterns among benign and malignant BCs, researchers used neural networks and ultrasound imaging with multi-fractal dimension characteristics. The majority of the work was done on imaging. The study reported the most precise classification results. Through the literature review, different machine learning approaches best suited for the objectives are evaluated since not many deep learning-based Raman spectral signal processing was done. Many cancers related studies employed several machine learning models, conducted on various subclasses of their data. In particular, much research compared the effectiveness of various ML models, typically conventional ML models including LDA and PCA with DL models such as CNNs [8]. To differentiate malignant BCs from benign, researchers used neural networks and ultrasound imaging with multi-fractal dimension characteristics. The majority of the work was done on imaging. The classification results of the study were precise with 82.04 percent of accuracy [17]. In the Weka tool for the identification of BCs, a comparative analysis of clustering approaches, such as LVQ, hierarchical clustering, DBSCAN, and canopy, was undertaken [18]. The first clustering technique was found to have comparatively the highest accuracy of 72 percent for prediction [18], according to the published results. For BC classification, CNN and Multiple Instance Learning (MIL) were coupled. The studies used the BreaKHis database, which included 8,000 malignant and benign BCs biopsy images. With a magnification factor of 40x, the classification rate was found to be 92.1 percent [19].

RNN is a sort of neural network (NN) model that can process both parallel and sequential data. By adding memory cells into the neural layer, similar operations to those of the human brain can be replicated. Bidirectional Recurrent Neural Network (BiRNN) is another RNN model that is modeled to process input sequences with known start and end in advance. Subsequently, RNN uses data from the previous session, and BiRNN might be used to make even more improvements. Data from 2 different sources can be handled by Bi-RNN. The sequence is processed by RNN from start to finish, while the other processes it backward from beginning to end [20], taking into account both the start and end context of each sequence element. LSTM-RNN and GRU-RNN are 2 RNN alternatives that differ in their gating units. The LSTM is a type of RNN that incorporates prediction based on context, which is not taken into account by regular RNNs. By training RNN, the vanishing gradients problem can be eliminated by LSTM.

1D-CNNs are also highly useful in Raman spectral classification. Hong *et al.*, for instance, performed a simpler network altered from the LeNet-5 architects with 2 convolutional layers (CL) only for the retrieval of features tailed by a fully-connected (FC) layer for the purpose of classification, which attained 96.33 % of accuracy [21]; to identify cancer in the prostate, Lenferink *et al.*, introduced 1D-CNN for Raman spectra for extracellular vesicles (EVs) [22]; and to evaluate the action of bowel disease. Furthermore, the accuracy for distinguishing prostate cancer, carcinoma in the lymph node of the prostate, red blood cells (RBC), and platelet was 93 %. Raffiee *et al.*, employed a 1D-CNN to identify microbial load by differentiating Raman spectra of Chinese hamster ovary cells from 12 kinds of microbes, achieving 95 - 100 % of accuracy [23]. To quickly identify components, Kalasuwan *et al.*, built a 1D-CNN for low-resolution Raman spectra recorded from BaSO4, NaNO3, Ba(NO3)2, Pb(NO3)2, CH4N2O, and KNO3 with an accuracy of 96.7 % [24]. This model is composed of 4 conv-blocks for the extraction of features and 1 output layer for the classification of the spectrogram. Aside from the aforementioned, Chen *et al.*, deployed a 1D-CNN to a nanoplasmonics biosensing chip (NBC), on a validation dataset which was accurately identified with an accuracy of 91 % [25]. This model has 8 kernels or 16 kernels of size 3×1 in 2 convolutional layers, correspondingly. Each convolution is preceded by a batch normalization layer, and the first convolution is followed by a 2-by-1 max-pooling layer. Finally, the results of classification are generated using a concatenate layer, an FC layer, and a softmax function.

In addition to the 1D-CNN techniques mentioned above, ResNets and autoencoders are often used for Raman spectral categorization. Glavin *et al.*, for instance, integrated an autoencoder with a locally connected neural network (LCNN) to construct a 2-step classifier model that is robust in the occurrence of negative anomalies [26]. The LCNN was built for training the dataset in this network, whereas the autoencoder was used for the detection of outliers. Jean *et al.*, developed 1 ResNet model for quick bacteria detection [27]. Their model's appropriate antibiotic detection accuracies were 97.0, 0.3 %. Furthermore, Zhu *et al.*, developed a new architecture called "diverse spectral band-based deep residual network", which had the optimum performance in identifying tongue squamous cell carcinoma with a 97.38 % of sensitivity, 98.75 % of specificity, and 98.25 % of accuracy, respectively [28]. Wang *et al.*, have created a new framework employing residual blocks called "multi-feature fusion convolutional neural network" (MCNN), which achieved the highest accuracy amongst other classifiers for thyroid dysfunction detection using Raman spectra of serum samples from 199 individuals [29].

For early-stage breast cancer identification and classification from blood plasma utilizing both Raman spectroscopy and deep learning, not many works are being conducted effectively, Krishna *et al.*, [30] and Bergholt *et al.*, [31]. employed RS to analyze the oral sample in an *in-vivo* environment. The findings of

the experiment indicate that the features of the spectral patterns are used to classify normal tissue specifically in high wavenumber. Singh *et al.*, put forwarded the implication of RS in oral cavity lesions detection [32]. RS has significant benefits in the detection of oral cancer tissue. The tumor cells alter the morphology and structure of tissues, and the resultant Raman peaks correlating to nucleic acid and proteins are evident [33]. Raman spectroscopy has significantly high specificity for the recognition of tumor tissue. It is hence condensed the necessity for adjunct therapy [34]. **Table 1** includes the model with comparatively the best performance performed on machine learning models, this helps to assist in the development of hybrid deep learning architecture for the analysis of Raman spectral signal against cancer markers. The metric is provided unless otherwise mentioned, despite the fact that its applicability to classification tasks is questioned, due to its ubiquitousness in the literature reviewed and its intuitive interpretation.

Table 1 Literature review: Overview of ML and DL techniques for spectral marker identification relating to cancer.

Authors/Year	Cancer sample	Model	Validation strategy	No. of spectra	accuracy
Zuvela, 2019 [35]	Tissue-Nasopharyngeal	GA- PLS-LDA	LOOCV	2,126	98 %
Yu, 2019 [36]	Tissue-Tongue	CNN	5-fold CV	1,440	97 %
Xia, 2020 [37]	Tissue-Tongue	CNN-SVM	5-fold CV	At least 216	99.5 %
Shu, 2021 [38]	Tissue-Nasopharyngeal	CNN	10-fold Venetian Blind CV	15,354	84 %
Sciortino, 2021 [39]	Tissue-Glioma	SVM	LOOCV	2,073	86 %
Riva, 2021 [40]	Tissue-Glioma	GB	LOOCV	3,450	83 %
KunxiangLiu, 2022 [42]	Gastric-Cancer Cell	SVM, KNN, LDA, XGBoost and DT	5-fold CV	660	99.20 %
Mariana Sarai Silva-López, 2022 [43]	Cervical Exfoliated Cells	PERMANOVA	CAP-cross-validation	81 subjects	90 %
Zirui Liu, 2022 [44]	Tissue-Gastric Cancer	SVM	20-fold CV	-	90 %
Liping Huang, 2023 [45]	Tissue-Hepatopathy	VGG-16	Cross-Entropy Loss	12,000	92.6 %
Zirui Liu, 2022 [44]	Tissue-Blood-Gastric Cancer	SVM	20-fold CV	-	72 %
Mehta, 2018 [41]	Brain Meningioma (serum)	PCA-LDA	LOOCV + independent test set	~8/subject	86 %
Lee, 2019 [22]	Tissue-Prostate	CNN	Single split 70/15/15	1,200	97 %
Jeng, 2019 [46]	Tissue-Oral	PCA-QDA	k-fold CV and LOOCV	400	82 %
He, 2021 [47]	Tissue-Renal	SVM	LOOCV	4,860	92.89 %
Chen, 2021 [48]	Tissue-Lung cancer & glioma	CNN	5-fold CV	2,700	99 %
Bury, 2019 [49]	Tissue-Brain Metastases	PCA-LDA	Not Stated	525	80.2 %
Aubertin, 2018 [50]	Tissue-Prostate	ANN	LOOCV	928	86 %
Yan, 2020 [51]	Tissue-Tongue	CNN ensemble	5-fold CV	At least 216	99 %
Wu, 2021 [52]	Tissue-Colon	CNN	LOOCV	2,420	94 %
Santos, 2018 [53]	Tissue-Skin	PCA-LDA	Single split 60/40	9-19/ sample	62.5 %

Many studies used different ML models on tissue samples and rarely on blood samples, which analyzed various subclasses of their spectral data and compared several other methods, generating a multitude of classification results. In addition, several studies compared the performance of different ML

models, often conventional ML models such as LDA against deep learning models; in particular using CNN with the Raman spectra obtained from tissues having different morphology. In most cases, the accuracy of a model was considered as the primary reported performance metric. Many ML and DL models perform better with the respective spectral data. Also, it is analyzed from the literature that the number of samples utilized was comparatively less for ML and deep learning models for evaluating the reliability of the models by considering the current clinical demands of Raman spectroscopy as a primary technique for biological sample analysis and assessment. For deep learning models the resultant accuracy was the no. of correct classifications divided by the total no. of classification attempts, though most of the models perform better, still face ambiguity in the classification accuracy due to the high dimensional Raman spectral patterns. This classification ambiguity is resolved by proposing hybrid deep learning models for training the Raman spectra. The objective of using hybrid deep learning models is to strengthen the role of Raman spectral data analysis to uplift the capability and standard of Raman spectroscopy in early-stage cancer diagnosis. It is also evident from the recent studies that not many kinds of research were conducted by employing the scope of using Raman spectroscopy and hybrid deep learning techniques for analyzing cancer from blood plasma samples.

Materials and methods

In this section, the materials and methods used for acquiring Raman spectral signals from histopathologically labeled blood plasma samples are discussed, which includes plasma collection and preparation of samples, Raman spectral acquisition, and analysis of characteristic features, preprocessing of Raman spectral signals to highlight the features relevant for further analysis and Raman spectral data augmentation for improving breast cancer classification accuracy.

Preparation of sample and protocol

18 females diagnosed with BC histologically and 8 control females were chosen for the study. Particularly, the available BC samples belong to stages 2, 3 and 4. Since candidates rarely report to the clinic at stage 1, stage 1 samples are not available. Candidates from similar socioeconomic and ethnic backgrounds were grouped. To get plasma samples, blood samples were taken in EDTA vials and centrifuged at the rate of 3,400 rpm for 5 min.

Acquisition of Raman spectral data

20 µl of a plasma sample from whole blood was placed on a substrate (aluminum) and spectral characteristics of Raman were obtained. 20 µl of the same plasma sample are placed on the aluminum substrate 3 times, with a new aluminum substrate each time, yielding a total of 25 spectra from each sample of blood plasma. Raman spectra were collected from all 18 females with BC and 8 controlled blood samples using a Raman micro spectrometer. The spectrometer has a laser source of 785 nm diode with 60 mW power. The Raman spectra were obtained between 600 and 1,800 cm^{-1} in 30 s. **Table 2** lists the Raman spectral feature assignments [54-58].

Table 2 Raman spectral feature assignments.

Spectral wavenumbers	Assignment names (Biomarkers)
625 cm^{-1}	Nucleotide conformation
640 cm^{-1}	Stretching of C-S & twisting of C-C in $C_9H_{11}NO_3$
675 cm^{-1}	CO-NH I (b-sheet)
689 cm^{-1}	Nucleotide conformation
714 cm^{-1}	C-N adenine CN_2 $(CH3)_3$ (lipids)
752 cm^{-1}	C_5 H_5 N_5 O
761 cm^{-1}	C_{11} H_{12} N_2 O_2, d (ring)
770 cm^{-1}	PO_4^{3-}
778 cm^{-1}	C_{47} H_{83} O_{13} P
788 cm^{-1}	phosphodiester bands in DNA

Spectral wavenumbers	Assignment names (Biomarkers)
798 cm^{-1}	CH deformation
828 cm^{-1}	$C_9H_{11}NO_3$ /protein
857 cm^{-1}	$C_{57}H_{91}N_{19}O_{16}$
885 cm^{-1}	Disaccharide (cellobiose), (C-O -C) skeletal mode
901 cm^{-1}	Monosaccharides (b-glucose)
917 cm^{-1}	C_5 H_9 NO_2, C_5 H_9 NO_3
932 cm^{-1}	C-C, a -helix
1,100 cm^{-1}	C-C gauche- bonded chain
1,138 cm^{-1}	n (C-C)- fatty acids, lipids
1,145 cm^{-1}	n (C-C)- fatty acids, lipids
1,173 cm^{-1}	$C_4H_5N_3O$, $C_5H_5N_5O$
1,185 cm^{-1}	Anti-symmetric PO_4^{3-} vibrations
1,268 cm^{-1}	d (=C-H) (CH_2OH-CHOH-CH_2OH)
1,285 cm^{-1}	CH_2OH-CHOH-CH_2OH
1,307 cm^{-1}	Twisting of CH_3/CH_2, wagging &/or $C_{65}H_{102}N_{18}O_{21}$ bending mode & $CH_3(CH_2)$ nCOOH
1,319 cm^{-1}	$C_5H_5N_5O$
1,410 cm^{-1}	ns COO_2 (IgG)
1,440 cm^{-1}	deformation of CH_2 in normal breast tissue
1,609 cm^{-1}	$C_4H_5N_3O$
1,660 cm^{-1}	$C_{15}H_{31}N_3O_{13}P_2$

Preprocessing of data

The objective of pre-processing is to highlight the Raman spectral and its characteristic features relevant for breast cancer. All Raman spectral data processing was done with Python [59]. Smoothing, vector normalization, substrate removal, and baseline correction were all part of the data pre-processing. The Savitzky-Golay smoothing method was used to vector normalize and smooth all spectra, including substrate backgrounds (order 5,13-point window). All of the spectra were rubber band corrected, and the spectra from the substrate were deducted from every spectrum to remove the baseline. Smoothing of the spectral data is a typical pre-processing technique. Smoothing is a mathematical process done on raw spectral data to eliminate (random) noise. This is particularly necessary when attempting to isolate significant spectral characteristic features that may be partially occluded by noise. For Raman spectroscopy, derivatives of the real spectral data are evaluated. The numerical distinction of data intensifies noise highly, and smoothing is required to obtain meaningful and significant spectral derivatives. The Savitzky-Golay (SG) method is a standard in data smoothing, especially in the case of time series data. Rubber band corrections are performed with the ends attached to the ends of the spectrum or at least 1 component of the spectrum to be corrected, and wrapped around the curve profile of the spectrum from below. The rubber band is then positioned against the spectrum's curve characteristic pattern. These are used in all deep learning approaches and multivariate analyses for time series data. When this is subtracted from the spectrum, the required spectrum with a rectified baseline is attained, and the preprocessing results are shown in **Figure 1**, where the spectral sample from cancer and controlled sample are preprocessed and the resultant results are shown in **Figure 1**.

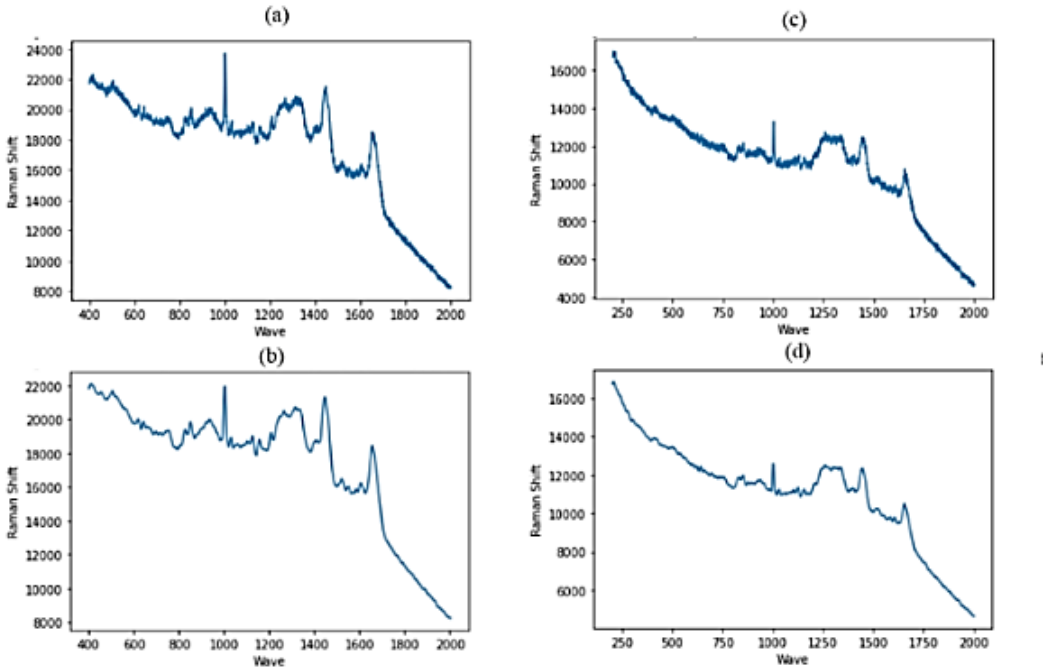


Figure 1 (a) the raw spectrum obtained from a BC sample, (b) the preprocessed spectrum of the raw BC sample, (c) the raw spectrum obtained form-controlled sample, and (d) the preprocessed spectrum of the raw controlled sample.

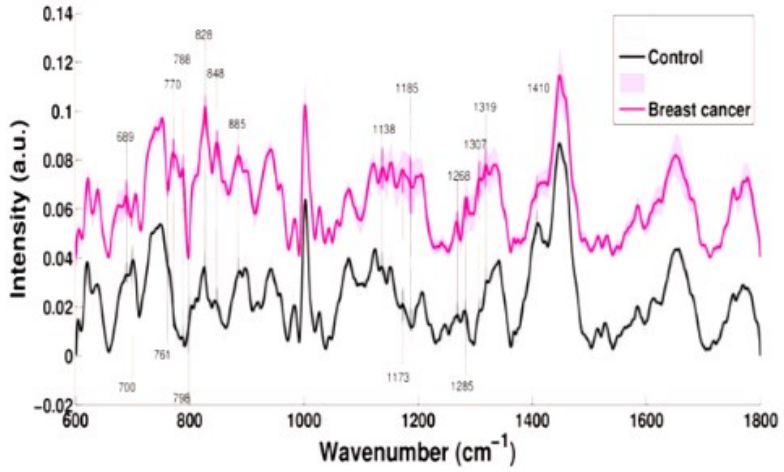


Figure 2 Healthy (black) and BC (magenta) blood plasma samples' mean Raman spectra.

Figure 2 illustrates the peaks of mean spectra of healthy and BC plasma samples against the markers listed in **Table 2**, with possible distinguishing features specified by vertical lines. Raman spectral features in the mean Raman spectra of patient samples, at 689, 770, 788, 828, 848, 885, 1,138, 1,173, 1,185, 1,268, 1,285, 1,307 and 1,319 cm^{-1} have intensities significantly higher. In the mean Raman spectra of healthy candidates, however, the features of Raman spectra at 700, 761, 798 and 1,410 cm^{-1} had higher intensities than in the patient samples.

Augmentation of spectral data

The insufficiency of Raman spectral data for the training of DL models for cancer diagnosis should be addressed along with overfitting, due to the patient's privacy concerns and other related limitations, very few reliable spectral samples are accessible for training the DL architectures. Often a small set of spectral data are being used for both training and testing of the algorithm. Based on the size of the spatial context and sampling scheme used in inputs, models can lead to certain overfitting and misleading high

classification accuracies on the spectral test dataset. Nevertheless, model overfitting and poor performance are common problems in applying neural network techniques [60]. In overfitting, the classification model attempts to learn too many characteristic details from the training of spectral data along with the artifact and noise. Hence, the performance of the model is considerably very poor on blind or test datasets. Therefore, the network fails to generalize the features or patterns present in the training dataset. To overcome these issues, deep learning models need sufficient data to process. This research also aims to propose solutions to tackle difficulties, this research introduces a generative adversarial network (GAN) to generate augmented Raman spectral samples. Adequate training of data allows the neural network to properly train the intrinsic features of the spectral data, improve the steadiness of models, and minimize overfitting. Hence data augmentation is important to lower model overfitting and help build a more reliable neural network model. Data augmentation prevents overfitting and enhances the learning capability. Though the number of observations (spectral sample) made for the study was limited, a successful and simple spectral data augmentation technique was created to synthesize new spectral data with no incurring additional labeling requirements. Data augmentation is a process that entails adding repeated spectra to the data by replicating along with noise and/or other changes to enhance the no. spectra. The data augmentation technique is employed in DL to not only enhance the spectral data size but also to add noise into the spectral data so that the tendency of the model to likely get overfit against the training set is less significant. As a result, the data is regularized, and the learning process is smoothed out. This technique doesn't need to confront biological proceedings, nonetheless is convenient as it can be used as a technique for data regularization, hence it reduces the tendency of a model to over-fit the data. **Figure 3** shows the proposed framework for RS-based spectral augmentation using the generative adversarial network.

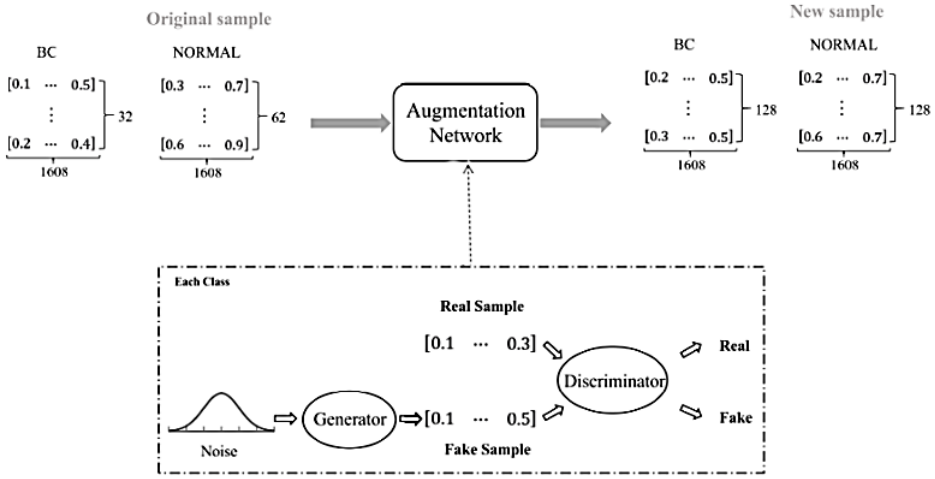


Figure 3 Data augmentation based on GAN for the generation of Raman spectral sample.

Algorithm for spectral data-augmentation (DA)
Input: $p_h(a)$: Noise prior to GAN training
C={c^1,…,c^n}: Original data of respective categorical classes
M: # for training GAN Des
Output: Spectral data augmentation
a. For no. training repetition DO
b. for m step do DO
c. Sampling of n noise {a^1,…,a^n}from noise prior $p_h(a)$;
d. Update the Des by maximizing the Eq. (1);
e. End
f. Sampling of n noise {a^1,…,a^n} from noise prior $p_h(a)$;
g. Update the Gen by minimalizing the Eq. (2);
h. end

Description of spectral data augmentation algorithm

Data augmentation algorithm illustrates the Raman spectral data augmentation module. A noise is given prior p_h (a) and original spectral data of respective class C = $\{c^1, ..., c^n\}$, the objective of this DA approach is to create new spectral samples for respective classes of original spectral data. At first, the process is alternated between *m* steps of optimizing *Des* and 1 step of optimizing *Gen*. Maximization of *Des* (Step c - d) loss-function. After *m* steps, minimalize the *Gen* (Step f - g) loss-function. The training set is enhanced for classification by adding the newly generated spectral samples to the original spectral dataset after the training process.

Spectral data augmentation

To examine the effect of spectral DA on the performance of the classifier, 2 spectral augmentation methods are provided: Balanced-DA and stratified-DA, to determine how augmented data should be incorporated for developing effective prediction models.

1) Balanced data augmentation (SDA_b)

This method aims to provide the same no. of generated samples to every class. For each class, n′ new samples are generated via data augmentation (there are 1,800 dimensions on each sample.). As a result, the no. of BC and normal samples increases to n' + 36, and n'+ 63, correspondingly. Following that, all spectral data are pooled to create a training spectral dataset of 2 * n' + 99 spectral data samples.

2) Stratified data-augmentation (SDA_s)

This method aims to keep the same class prior probability throughout the process of spectral DA, by creating various no. of augmented spectral samples for respective classes according to the proportion *m* between various classes of the original spectral data. Consequently, the no. of augmented spectral samples is m * 36 for BC and m * 63 for control.

To address the issue of insufficient spectral sample size, GAN is used to explore data augmentation. GAN has been used in various research to generate high-quality data for incorporating training data sets [61]. The GAN is comprised of 2 major components, as represented in **Figure 3** by the dashed rectangular box: Discriminator (Des) and generator (Gen). Both blocks are made up of multi-layer perceptron. The Gen is responsible for creating identical fake spectral samples from the latent vector *z*. The Gen is regarded as a group of counterfeiters attempting to create fake identical spectral samples to persuade the *Des* to assign the generated spectral samples a higher score.

The *Des* attempts to distinguish between the original spectral data samples and the generated spectral data samples. The *Des* and *Gen* are running in an adversarial way to enhance each other. The *Des*, in particular, attempts to understand the original spectral data and leads the *Gen* by delivering feedback on the generated augmented spectral samples. *Gen* learns from the feedback and attempts to generate new spectral samples that are very close to the original data. The *Des* is trained to maximize the likelihood of differentiating between original spectral samples and samples created by the *Gen*. The *Des's* loss-function is expressed as follows:

$$max_D V(Gen, Des) = E_{x\sim p_d}[logDes(x)] + E_{x\sim p_z}[\log(1 - Des(Gen(z)))] \quad (1)$$

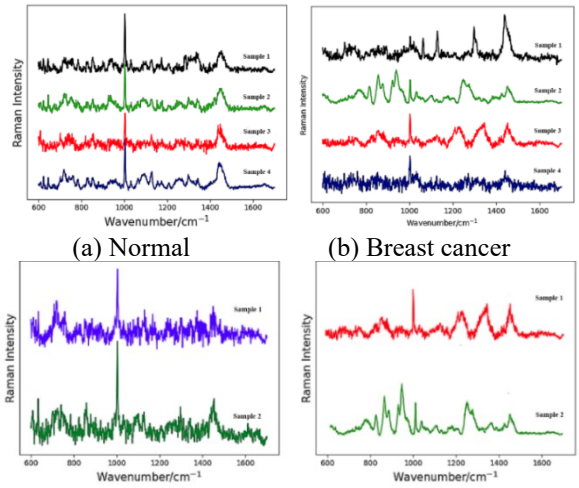
Concurrently, the *Gen* is trained to minimize log (1 *Des* (*Gen* (z))), to generate spectral samples as similar to real training spectral samples as possible. *Gen* loss-function is expressed as follows:

$$min_D V(Gen, Des) = E_{x\sim p_z}[\log(1 - Des(Gen(z)))] \quad (2)$$

Des and *Gen* are trained using the loss function from 2-player minimax:

$$max_D min_D V(Gen, Des) = E_{x\sim p_d}[logDes(x)] + E_{x\sim p_z}[logDes(Gen(z)))] \quad (3)$$

A new set of spectral samples are acquired for each class of the original spectral data from the data augmentation technique, which will be utilized to improve the classifier's performance. **Figures 4(c) - 4(d)** visualizes augmented Raman spectral samples created using GAN, pertaining to each class of samples. It is shown that the spectral samples created from GAN are very identical to the actual raw Raman spectral data seen in the a and b, which proves the efficiency of the spectral data augmentation. This means that the augmented data are not only visually identical but also reserve identical spectral features distribution, as raw genuine samples.



(a) Normal (b) Breast cancer

(c) Augmented Normal (d) Augmented breast cancer

Figure 4 Real raw Raman spectra (a - b) obtained from a dataset with 2 classes: BC and normal. Each sample contains 1,608 dimensions, with each dimension corresponding to the wavelength no. of the Raman shift (range from 600 to 1,700 cm^{-1}). The intensity of Raman is represented by the value of each dimension. Each colored curve refers to 1 Raman spectral sample. Rows (c - d) show GAN-generated augmented Raman spectral samples for each class (BC and normal). Each class has 2 Raman spectral samples generated.

Methodology

The aim of the proposed research is to apply various hybrid DL techniques to effectively determine whether a patient has BC or not from the Raman spectrogram. It is seen from various other studies discussed in the literature review section that deep learning techniques perform better than normal conventional machine learning architecture for different applications. Hence it is evident that the application of hybrid deep learning by combining the qualities of different deep learning models yields better results. Here the Raman spectral data are processed with Hybrid Deep learning approaches for automatic recognition of features from raw spectral data associated with cancerous samples with an end-to-end training procedure archetype. The Hybrid DL techniques are used to classify blood samples and differentiate whether they are malignant or benign. For such predictions, the put forth method uses ensemble 1D-CNN–LSTM,1D-CNN-GRU, CNN-BiLSTM, GRU-LSTM, RNN-LSTM, and OGRU-LSTM framework. The potential of these hybrid architectures is put forward in this research work, which has a different pattern of CNN, LSTM, and GRU layers. The validations of the HDL models are performed by validating the throughputs obtained from the classifiers and analyzing with the ground truth values which it is histopathologically labeled for each sample. The results of each HDL classifier are cross-validated with the label marked against each sample histologically to ensure classification performance. **Figure 5** shows the workflow diagram of the proposed HDL methods.

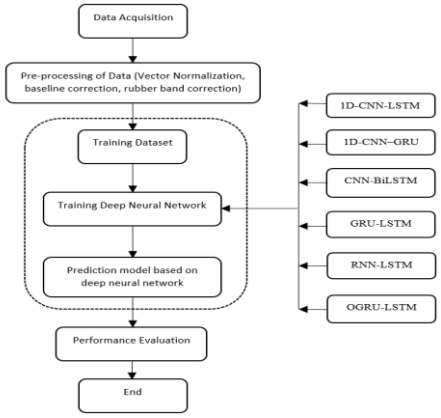


Figure 5 The flow diagram of the prediction model.

Overview of deep learning

Raman spectroscopy has been constantly moving to explore new possibilities, with the help of quantitative data analysis. It enables researchers to adapt spectral interpretations and get beyond the Raman instrument's physical constraints. In recent years, chemical researchers have paid increased attention to deep learning, a potent machine-learning technique for creating predictive and exploratory models from large spectral raw datasets. DL is a type of representation learning that can resolve a variety of AI challenges [62]. DL models are particularly useful in Raman spectroscopy for internal data pre-processing of spectral data and modeling of the same, which can potentially be used for all types of Raman spectral data. In this aspect, deep learning models are a prerequisite for Raman spectroscopy for both data processing and data modeling using higher dimensional spectral features. When a large no. of Raman spectra is available, they are directly sent into deep learning models with no prior processing. In Raman spectral data analysis, the total volume of data plays a decisive role. DL methods give better performance for higher dimensional data sets. Hence the scope of hybrid models by combining different good-performing deep learning models increase the accuracy for the spectral pattern analysis and binary classification of higher dimensional Raman spectral data compared with single models on breast cancer spectral datasets, especially the combination of deep learning models with GRU, 1D-CNN, RNN and so on. Hybrid Deep Learning, which combines different DL algorithms, has gained momentum over the last few years. As a result, the next sections will delve into greater detail about hybrid deep learning algorithms, their relevant uses in Raman spectroscopy especially in breast cancer diagnosis, and their current problems. Combining 2 or more deep learning classifiers leads to the formation of a hybrid deep neural network model by modulating the neural network's weight initialization and the input feature diversity.

Model selection for HDL

The selection of models is based on several reason depending on the characteristics of the deep learning architectures for processing the time series spectral data. The deep learning models especially OGRU [67], LSTM [64], 1D CNN [65], GRU [63], BiLSTM [68], RNN [69], and CNN [66] are specifically designed to process sequential data, and are able to capture complex dependencies and patterns over time in spectrogram. This makes them well-suited for modeling Raman spectral patterns. These models have achieved state-of-the-art performance on a variety of time series tasks, including forecasting, anomaly detection, and prediction of complex systems and are scaled to handle large datasets. These models can be applied to a wide range of time series data, including data with long-term dependencies, high-frequency data, and data with complex patterns. Hence combining different deep learning models can be beneficial for time series data analysis since it leverages the strengths of each model while mitigating their weaknesses in Raman spectral signal processing. For instance, the combinations of selected deep learning models; OGRU and LSTM models are well-suited for modeling long-term dependencies in breast cancer spectral data, while 1D CNN and GRU models are effective at extracting features from the Raman spectral data. BiLSTM models can capture dependencies in both directions, while RNN and CNN models are able to process sequential data. By combining these models, it is able to achieve better performance and increased robustness compared to using a single deep learning model. Additionally, combining multiple models can reduce the risk of overfitting and provide increased interpretability of the underlying patterns and relationships in the Raman spectral data.

OGRU- LSTM model

This section details the hybrid model for breast cancer identification and classification and it is seen that the same model can enhance the accuracy of classification. Considering all the potential advantages of hybrid models and focusing on enhancing the performance of classification approaches, this paper compares and evaluates the put-forward hybrid model with the other 3 leading deep learning models. These methodologies are applied to classify the Raman spectrum as either benign or malignant.

GRU, on the other hand, has drawbacks such as low learning efficiency and slow convergence. To eliminate these drawbacks a neural network model with an OGRU is proposed. The OGRU architecture utilizes the reset gate to advance the prediction performance and learning efficiency of GRU by optimizing its learning process. As a result, an OGRU is proposed, since OGRU is derived from the GRU architecture [67], in which the GRU neural unit's update gate is enhanced, the actual input of update gate e_t is modified to e_t multiplied with r_t, and the output of the reset gate is utilized to adjust the update gate with feedback. By filtering the present input data e_t through the reset-gate, similar data's negative effects are reduced to a greater extent, and convergence will be accelerated, hence the goal of efficient learning will be achieved. The deep OGRU neural network is modeled from GRU architecture, whose neural structure is depicted in **Figure 6**.

Consider $e_1, e_2..., e_t$ as the input sequence, then the gate is updated at t, reset the gate, and apply OGRU output calculation formula:

$$q_t = \sigma\,(W_q * [r_{t-1}, e_t]) \tag{4}$$

$$l_t = \sigma\,(W_l * [r_{t-1}, e_t * r_t]) \tag{5}$$

$$m_t = tanh\,(W * [q_t, r_{t-1} * e_t]) \tag{6}$$

$$r_t = (1 - l_t) * r_{t-1} + l_t * m_t \tag{7}$$

$$k_t = \sigma\,(W_o * r_t) \tag{8}$$

Among them, In the formula, r_t and l_t have the same denotation as normal GRU neurons. The OGRU neuron varies from the GRU neuron in that, in order to hide the state weight, the r_t is multiplied with the preceding time at the update gate l_t, allowing the reset gate to re-screen the present input data e_t, and to optimize the neuron structure, the reset gate's output is used to modify the update gate.

The neuron topology of the OGRU is more logical than GRU, as shown in **Figure 6** and formula (5), and at each time the hidden state can be made simpler, and to some extent, the gradient attenuation can be controlled. As a result, the OGRU model may retain a higher distance information dependency, as well as a higher prediction accuracy and learning efficiency [67].

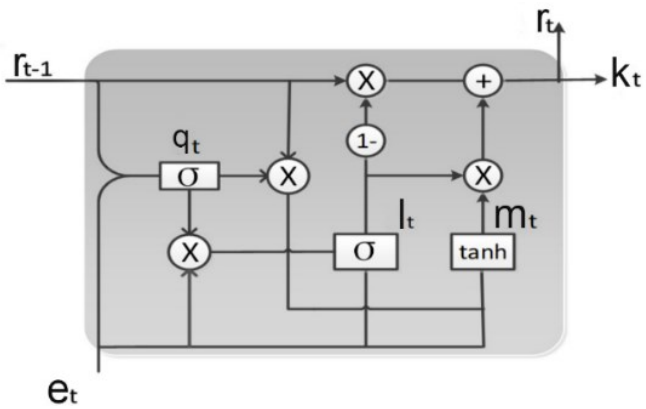


Figure 6 The OGRU neuronal structure [67].

The input, output, and hidden layers make up the deep OGRU-NN. The OGRU neurons constitute the hidden layers. The recursive transfer of data among neurons is increased, and the data preservation capability is enhanced, by optimizing the learning process of the GRU model. It is significant that the ratio of the mean of error and the sum of squares of error are employed to evaluate the accuracy of prediction. They are commonly used indicators for determining the variation between actual and expected values, and they can correctly reflect the model's accuracy of prediction [67]. Formulas (9) for AE and (10) for SSE illustrate the error ratio means and the sum of error squares:

$$AE = \frac{1}{M}\sum_{i=t}^{M}\left|\frac{k_i' - k_i}{k_i}\right| \tag{9}$$

$$SSE = \sum_{i=1}^{M}(k_i' - k_i)^2 \tag{10}$$

M denotes total no. samples in the validation, k_i' are the output predicted, and k_i is the real output. The accuracy of prediction is higher if AE is smaller than SSE.

Table 3 Proposed stacked ensemble OGRU-LSTM model.

Name of layers	Dropout rate/ No. of nodes	Shape of output	No. of parameters received
Bi-directional_OGRU	512	Nil-30-124	99,850
Layer-Dropout	0.2	Nil-30-512	0
Bi-directional_OGRU	256	Nil-30-256	164,362
Layer-Dropout	0.2	Nil-30-128	0
Bi-directional_OGRU	128	Nil-30-64	30,922
Layer-Dropout	0.2	Nil-30-64	0
Bi-directional_OGRU	64	Nil-32	20,396
Layer-Dropout	0.2	Nil-32	0
Layer LSTM	32	Nil-32	10,474
Layer Dropout	0.2	Nil-32	0
Layer LSTM	16	Nil-32	5,688
Layer Dropout	0.2	Nil-32	
Layer Dense	8	Nil-8	274
Layer Dense	4	Nil-4	46
Layer Dense	2	Nil-2	20
Layer Dense	1	Nil-1	4

To obtain maximum efficacy, hyperparameters must be tuned while designing this model. This section describes the model's specifications as well as its hyper-parameters. By considering the ability of LSTM with its less memory consumption, and fast and accurate performance in using datasets with longer sequences as pertaining to this work's spectral data, the OGRU model is ensembled with LSTM. The model proposed has 4 OGRU and 2 LSTM layers, each with 512, 256, 128, 64, 32 and 16 units. Each of these layers are having a 20 % dropout rate. Following that, 4 Dense (D)-layers with 8, 4, 2 and 1 node are stacked in this model, correspondingly. With the exception of the dropout layers (DoL), in all levels, the sigmoid activation function is used. These layers are finally compiled using a 64-batch size Adam optimizer with 100 epochs. The hyper-parameters can be adjusted to help the model achieve the best prediction results. To obtain prediction, the NN takes a total parameter of 3, 11, 640 and trains them. **Table 3** has a comprehensive description of the model.

Stacked RNN-LSTM model

Each of the 4 LSTM layers is followed by dropout layers (DoL) in a stacked LSTM model. Four D-layers are included in the model again. The activation function 'sigmoid' is used to implement the final D-layer and the LSTM layers. This model trains with received 131,702 parameters for obtaining results prediction once it is implemented by selecting relevant hyper-parameters. **Table 4** describes this model.

Table 4 Stacked RNN-LSTM model.

Name of layers	Dropout rate/No. of nodes	Shape of output	No. of parameters received
Layer LSTM	128	Nil-30-128	66,560
Layer Dropout	0.2	Nil-30-128	0
Layer LSTM	64	Nil-30-64	49,408
Layer Dropout	0.2	Nil-30-64	0
Layer LSTM	32	Nil-30-32	12,416
Layer Dropout	0.2	Nil-30-32	0
Layer LSTM	16	Nil-16	3,136

Name of layers	Dropout rate/No. of nodes	Shape of output	No. of parameters received
Layer Dropout	0.2	Nil-16	0
Layer D	8	Nil-8	136
Layer D	4	Nil-4	36
Layer D	2	Nil-2	10
Layer D	1	Nil,1	3

Stacked GRU-LSTM Model

This model, like the other baseline classifiers, is made up of GRU sequences and dropout layers. This model similarly has 4 D-layers layered on top of each other. Like the previous 2 models, the final output layer and the GRU layers both take the activation function sigmoid. The model proposed has 4 GRU and 2 LSTM layers, each with 512, 256, 128, 64, 32 and 16 units. Each of these layers are having a 20 % dropout rate. Following that, 4 D-layers with 8, 4, 2 and 1 nodes are stacked in this model, correspondingly. With the exception of the DoL, in all levels, the sigmoid activation function is used. These layers are finally compiled using a 64-batch size Adam optimizer with 100 epochs. The hyper-parameters can be adjusted to help the model achieve the best prediction results. To obtain prediction, the NN takes a total parameter of 2, 62, 439 and trains them. **Table 5** gives a brief detail of the model.

Table 5 Stacked GRU-LSTM model's description.

Name of layers	Dropout rate/No. of nodes	Shape of output	No. of parameters received
Layer GRU	512	Nil-30-124	86,860
Layer Dropout	0.2	Nil-30-512	0
Layer GRU	256	Nil-30-256	135,461
Layer Dropout	0.2	Nil-30-128	0
Layer GRU	128	Nil-30-64	40,822
Layer Dropout	0.2	Nil-30-64	0
Layer GRU	64	Nil-32	22,373
Layer Dropout	0.2	Nil-32	0
Layer LSTM	32	Nil-32	9,586
Layer Dropout	0.2	Nil-32	0
Layer LSTM	16	Nil-32	7,698
Layer Dropout	0.2	Nil-32	0
Layer Dense	8	Nil-8	364
Layer Dense	4	Nil-4	53
Layer Dense	2	Nil-2	32
Layer Dense	1	Nil-1	12

CNN-BiLSTM-AT model

This section describes the architecture of the model employed in this study for CNN-BiLSTM-AT. By taking the parametric values of the spectral signal along with features as an ordered sequence, the BiLSTM has become an effective approach for modeling spectral pattern curves, since it can capture information from a data sequence as well propagate information from adjacent dependences. However, the benefits of CNN and the attention process in extracting characteristic features are discussed. **Figure 7** depicts the structure of the proposed hybrid CNN-BiLSTM-AT method. This architecture is divided into 2 branches, one of which uses CNN to record the attributes of spectral signal and the other which performs a selection of features using a 2-layer BiLSTM with an attention mechanism. To obtain the predictions, the

characteristic features from different branches are combined and put into the final dense layer. Eq. (11) can be used to express the specified sequence matrix *e*.

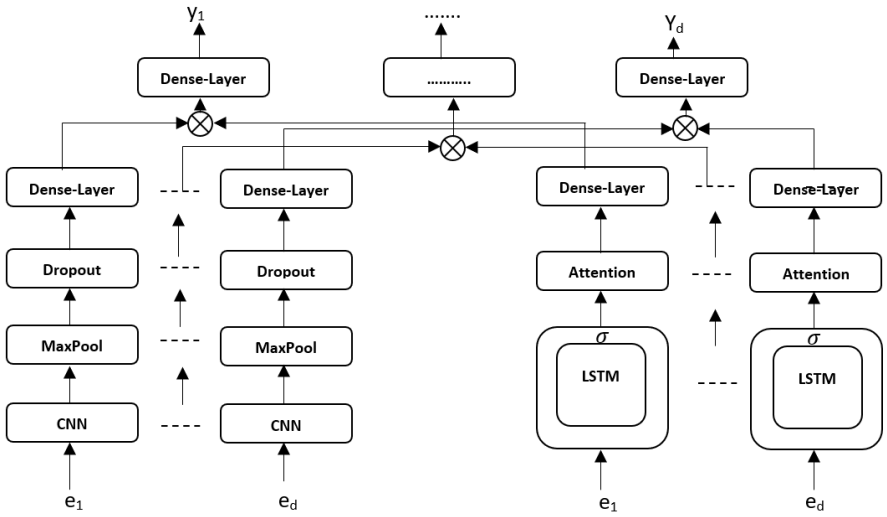


Figure 7 The architecture of CNN-BiLSTM-AT.

$$x = \begin{bmatrix} e_1^1 & e_1^2 & \dots & e_1^n \\ e_2^1 & e_2^2 & \dots & e_2^n \\ . & . & . & . \\ e_i^1 & e_i^2 & \dots & e_i^n \\ . & . & . & . \\ e_d^1 & e_d^2 & \dots & e_d^n \end{bmatrix} \quad (11)$$

wherein $e_d = (e^1_d, e^2_d, \dots, e^n_d)$ represents the parameters of *n* spectral signal at feature depth *d*. The set $e_n = (e^n_1, e^n_2, e^n_I \dots e^n_d)$ represents the values of a specific spectral feature curve. $Y = (y_1, y_2, \cdots, y_i, \dots, y_d)$ is the forecasted result. The X is fed to convolution operators in the CNN branch. To collect low-level implicit characteristics features from raw spectral signal, 1-D conv-operators glide 128 filters with the same size as window 1 over an ordered sequence. The dimension of characteristic feature maps is then reduced via a max-pooling function. To prevent the CNN from overfitting, a dropout function is used. ReLU is the non-linear activation function of this layer. Lastly, a flattened layer is utilized to reverse the direction of dimension, followed by a dense layer with 100 neurons to perform a linear operation. The two-layer BiLSTM captures contextual data from the original spectral signal in the BiLSTM-AT branch. Every BiLSTM with 50 cells not only learns from the previous term but also from the following term of the current sequences, so the characteristics and features captured by BiLSTM are considered as 2 alternative interpretations of the spectral signal. The BiLSTM features are sent into the attention layer and then appended to the output of the fully connected unit. One dense unit follows the attention layer. The CNN and BiLSTM-AT branch output is merged before being transmitted into the final dense layer with 1 neuron. Tanh is the dense layer's activation function. The CNN-BiLSTM-AT network's objective function is set to RMSE throughout the learning phase.

The model proposed has 4 CNN and 2 BiLSTM layers, each with 512, 256, 128, 64, 32 and 16 units. Each of these layers are having a 20 % dropout rate. Following that, 4 D-layers with 8, 4, 2 and 1 nodes are stacked in this model, correspondingly. With the exception of the DoL, in all levels, the sigmoid activation function is used. These layers are finally compiled using a 64-batch size Adam optimizer with 100 epochs. The hyper-parameters can be adjusted to help the model achieve the best prediction results. To obtain prediction, the NN takes a total parameter of 2, 57, 827 and trains them. **Table 5** gives a brief detail of the model.

Table 6 Structure of the CNN-BiLSTM model.

Name of layers	Dropout rate/No. of nodes	Shape of output	No. of parameters received
Layer-CNN	512	Nil-30-512	96,869
Layer-Dropout	0.2	Nil-30-512	0
Layer-CNN	256	Nil-30-256	121,034
Layer-Dropout	0.2	Nil-30-128	0
Layer-CNN	128	Nil-30-64	38,696
Layer-Dropout	0.2	Nil-30-64	0
Layer-CNN	64	Nil-32	21,587
Layer-Dropout	0.2	Nil-32	0
Layer-BiLSTM	32	Nil-32	10,368
Layer-Dropout	0.2	Nil-32	0
Layer-BiLSTM	16	Nil-16	6,987
Layer Dropout	0.2	Nil-16	0
Layer Dense	8	Nil-8	782
Layer Dense	4	Nil-4	146
Layer Dense	2	Nil-2	36
Layer Dense	1	Nil-1	18

1D-CNN-GRU model

A CNN has [72], great feature extraction capacity which enables it to retrieve spectral features. A CNN is used as a signal processing and 2-dimensional image processing archetype [73]. To excerpt spectral features and achieve spectral identification, a conv-kernel runs smoothly over the spectrogram. Since the data used to predict the remaining useful features of chemical characteristics is primarily 1-dimensional, these spectral features can't be obtained utilizing a 2-dimensional convolutional kernel. Hence, a 1D-CNN [74], architecture is used to acquire sensitive spectral information and obtain properties of molecular feature extraction in the feature extraction step, depending on the sequence data attributes. In contrast to conventional feature extraction methods, CNN may extract spectral features with little effort. Table 7 illustrates the parameters of the put-forward 1D-CNN-GRU model.

The ID-CNN-GRU network is composed of 12 layers to model the unified neural network: Three layers of the 1D convolutional neural network, 2 layers of GRU [75], 2 layers of dropout, 1 layer of max-pooling, a flatten layer, and finally 3 FC layers. The spectral signals are primarily directed through the first conv-layer. The kernel used in the first layer of the ID-CNN-GRU network is as 2, signifying that all weights are shared by all strides of the input layer of indexes 0 and 1, as well as the outputs of the scalar coupling constant. The result of the first conv-layer is 2×128, but it's also fed as the input to the next conv-layer. The size of the filter is reduced from 128 to 64 in the second layer of 1D-convolutional, and the kernel size is reduced to 3 for all 1D-conv-layer. These layers are finally compiled using a 64-batch size Adam optimizer with 100 epochs. The hyper-parameters can be adjusted to help the model achieve the best prediction results. To obtain prediction, the NN takes a total parameter of 2, 79, 962 and trains them. **Table 7** gives a brief detail of the model.

Table 7 Stacked 1D-CNN-GRU model.

Name of layers	Dropout rate/No. of nodes	Shape of output	No. of parameters received
Layer-Convolution1D		Nil-30-128	96,640
Layer-Convolution1D		Nil-30-128	0
Layer-Convolution1D		Nil-30-64	142,352
Layer-MaxPooling1D		Nil-30-32	0
Layer GRU	128	Nil-16	33,104
Layer Dropout	0.2	Nil-16	0
Layer GRU	64	Nil-16	7,684
Layer Dropout	0.2	Nil-16	0
Layer Flatten	none		
Layer Dense	512	Nil-8	136
Layer Dense	256	Nil-4	36
Layer Dense	1	Nil-2	10

1D-CNN-LSTM model

This model's architecture is identical to that of 1D-CNN-GRU. The only distinction is that the GRU layer is replaced with an LSTM layer. The 1D-CNN-LSTM model has 3, 09, 122 trainable parameters, which is greater than the 1D-CNN-GRU model. Four conv-layers, input-layer, 1 pooling-layer (PL), 4 fully-connected (FC) layers, a soft max output layer, and 2 LSTM-layers comprise the 1D-CNN-LSTM model. **Figure 8** depicts the detailed network structure. To initiate, the 1D spectral signal data is directly fed as the input to the model, and its shape is 178×1. The input spectral data is then allowed to pass through the first conv-layer to extract the characteristic features from the raw spectral signal, when each conv-kernel is 3×1, there are 64 1D conv-kernels in the Conv-Layer 1, and there is a 1 stride between conv-kernels. Following this conv-layer is a ReLU activation layer, which can initiate non-linearity into the proposed archetype.

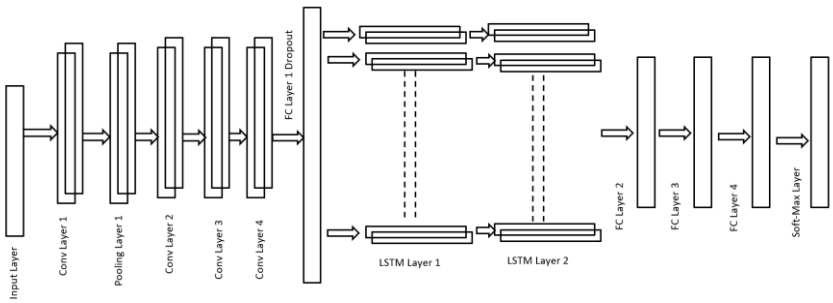


Figure 8 The architecture of the 1D-CNN-LSTM model.

The following gives the computational definitions of the 1D convolutional operation and ReLU activation.

$$y_j^l = \sigma\left(\sum_{i=1}^{N_{l-1}} conv1D\left(w_{i,j}^l, x_i^{l-1}\right) + b_j^l\right) \quad (12)$$

wherein $x^{l-1}{}_i$ indicates the feature map in i^{th} position in the layer $(l\text{-}1)^{th}$; the j^{th} feature map in the l^{th} layer is indicated by $y^l{}_j$; $w^l{}_{i,j}$ indicates the convolutional kernel to be trained; N_{l-1} indicates $(l\text{-}1)^{th}$ layer characteristics feature map; the 1D operation on convolution with no zero-padding is indicated by conv1D, hence, the feature map dimension in the l^{th} layer is lower than $l\text{-}1^{th}$ layer; the bias of the j^{th} feature map in the l^{th} layer is indicated by $b^l{}_j$; the ReLU activation function is indicated by σ(), ReLU helps avoid the problem of overfitting and is defined as below:

$$\sigma(x) = \begin{cases} o, x \leq 0 \\ x, x > 0 \end{cases} \quad (13)$$

The Pooling Layer1 pooling window size is 2 and the window stride is also 2. It has the potential to significantly lower the no. of training attributes in the proposed model while also accelerating the process of training. Dropout can mostly alleviate the difficulties of overfitting while FC Layer1 can combine the output from the conv-layers and lower the dimension of feature maps to fit the input layers of LSTM. The gates in LSTM can cooperate to retain prior information and enhance the capacity to learn useful information from Raman spectral data. Both the LSTM Layer 1 and the LSTM Layer 2 contain 64 neurons. The overview of this model configuration is summarized as shown in **Table 8**. These layers are finally compiled using a 64-batch size Adam optimizer with 100 epochs. The hyper-parameters can be adjusted to help the model achieve the best prediction results. To obtain prediction, the NN takes a total parameter of 3, 09, 122 and trains them.

Table 8 Stacked 1D-CNN-LSTM model.

Name of layers	Dropout rate/No. of nodes	Shape of output	No. of parameters received
Layer-Input	256	Nil-30-128	176,640
Layer-Conv1D	128	Nil-30-128	0
Layer-MaxPooling1D	2	Nil-30-64	93,352
Layer Dropout	0.2	Nil-30-32	0
Layer-Conv1D	128	Nil-16	31,084
Layer-MaxPooling1D	2	Nil-16	0
Layer Dropout	0.2	Nil-16	7,864
Layer LSTM	256	Nil-16	0
Layer Dropout	0.2		
Layer LSTM	32	Nil-8	136
Layer Dropout	0.2	Nil-4	36
Layer Flatten	none	Nil-2	10
Layer Dense	128		
Layer Dense	4		

Results and discussion

Hybrid deep learning models are being trained by feeding the dataset during the process of training. For testing and training, the resultant spectral dataset is employed. From the total of 26 blood samples around 1,950 Raman spectrums were obtained. When augmentation using GAN is performed on the obtained spectrums and thereby forms around 62,199 augmented spectrums with characteristic spectral features. With random sample selection, the data sets were divided into a training set, which included 3-quarters (3/4) of the spectra, and a validation set, which included the final quarter (1/4) of the spectra. Over each epoch, the process of training is assessed against accuracy and loss. Baseline classifiers are trained using epoch sizes of varying lengths. The initial epoch sizes were 20-epoch, 50-epoch, and 100-epoch. For training all the models, these epoch sizes were utilized.

Various hybrid deep neural network models are established in this study. The CNN-based architecture is responsible for collecting the spatial-spectral properties of the spectral signal [76], the GRU-based architecture is responsible for extracting the temporal spectral features [77], and the attention mechanism is employed to concentrate on improving the critical data [78]. To satisfy the target prediction, the Spatiotemporal connections from CNN and GRU network are combined with the attention mechanism along with LSTM. All loss functions can rapidly lower and become stable using the 1D-CNN-GRU technique, indicating that the models are trained well and can classify efficiently. The exploratory results show that the 1D-CNN-GRU technique can achieve a high degree of fitting correlation between prediction and actual spectral signal. The low computational cost and effectiveness of classifying spectral signals

based on the HDL technique make it possible to employ the same on large-scale spectral data like Raman spectral signals, which is conducive to the assessment and analysis of blood plasma sample characterization. Furthermore, by modifying the hyperparameters, trained models can be adapted for efficient prediction and classification. The modality of hyperparameters is the same as other conventional deep learning algorithms. The tuning hyperparameters are the no. of neurons, optimizer, no. of layers in activation function, rate of learning, epochs, and batch size. The accuracy of the model can be affected by different layers. Underfitting may be the result of fewer layers whilst overfitting is the result of too many layers. The no. of neurons in every hidden layer is considered the first hyperparameter to tune. Hence the no. of neurons in all layers of the models is set to be the same though it can be made altered. The no. of neurons is adjusted for the solution complexity. More neurons are required for the operation with a more difficult level of prediction. The range of the no. of neurons is set to 10 to 256. If the spectral size of the training sample is too large, it will considerably take an extended time to build the learning model. In order to learn the model faster, assigning a batch size hence not all of the data to be trained are fed to the model at the same time. The batch size is the total no. of input sub-samples of training data. If the batch size is 1,000 and there are 46,590 observations in the training dataset, the model will train 46 times using 1,000 training data sub-samples while the rest of the learning uses 590 training data sub-samples. As the size of the batch is smaller the process of learning is faster, but the variance of the validation dataset accuracy is higher. As the size of the batch is higher the process of learning is slower, but the accuracy of the validation dataset has a lower variance.

The training loss and validation loss of these HDL methods when they were applied to the spectral patterns of the scalar coupling constant are shown in **Figures 9(a) - 9(f)**. In terms of testing accuracy, the models are analyzed and compared after they had been trained. The highest predictive efficiency is listed in **Table 9** with an epoch size of 100. Henceforth, the optimal training criterion is an epoch size of 100. All of the models are executed using the Keras [70], deep learning framework with the TensorFlow [71] backend. **Figure 9(a)** shows that the model 1D-CNN-GRU validation and training losses were significantly reduced at the slowest speed. The validation and training loss scores of the 1D-CNN-GRU hybrid model were lower than those of the other HDL models, resulting in higher validation and training accuracies. Moreover, the performance of training and evaluation of the 1D-CNN-GRU was superior to that of the other models, and integrating GRU with 1D-CNN seems improved the performance significantly. Table 10 shows the performance analysis and errors for various models such as 1D-CNN-GRU, 1D-CNN-LSTM, CNN-BiLSTM, GRU-LSTM, RNN-LSTM, and OGRU-LSTM. **Table 10** shows that the metrics of RMSE, MAE, and MSE for the model 1D-CNN-GRU were all comparatively small. F1-Score, accuracy, and Cohen-Kappa Score of the 1D-CNN-GRU model are all assessed. Losses that occur during testing are also recorded. It is then compared against the models 1D-CNN-LSTM, GRU-LSTM, and CNN-BiLSTM, models. After the training procedure is completed, the accuracy of the test is calculated after the 100^{th} epoch. Using evaluation metrics, all the executed models are illustrated after the completion of 100 training epochs in terms of performance. Predictions for the test dataset are obtained after this training session utilizing training data. Such prediction results are compared to actual observable values, resulting in a deep model evaluation based on the metrics utilized. **Table 10** shows the results of the comparison analysis. The comparison analysis shows that the 1D-CNN-GRU model outperforms other put-forward hybrid deep learning classifiers in terms of promising results. The output attributes of the 1D-CNN-GRU model outperformed all other HDL models because of the union of 1D-CNN and GRU network attribute initialization.

Table 9 Accuracy of prediction for all different models of RNN for various size of epoch.

Size of epoch	RNN-LSTM	CNN-BiLSTM	GRU-LSTM	OGRU-LSTM	1D-CNN-LSTM	1D-CNN-GRU
20	67.73	67.79	67.73	67.79	69.79	83.67
50	70.72	78.76	70.72	78.76	83.94	92.76
150	84.63	87.01	89.27	92.63	95.32	98.90

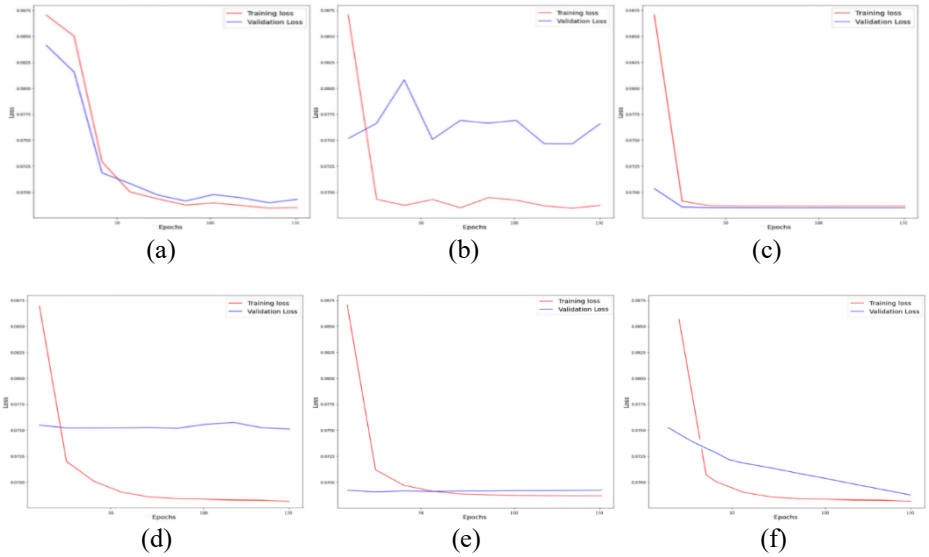


Figure 9 (a) Loss curve of 1D-CNN-GRU, (b) Loss curve of OGRU-LSTM, (c) Loss curve of 1D-CNN-LSTM, (d) Loss curve of CNN-BiLSTM, (e) Loss curve of GRU-LSTM, and (f) Loss curve of RNN-LSTM.

Training of models and evaluation of test performance

Pytorch and Sklearn were used in this work to implement 6 HDL methods for all training and testing operations. Simultaneously, the Pytorch was used to import the Data-loader class, and Shuffle was set to true as an attribute to assure that all data, including the BC and normal classes, were divided randomly into test-set and training-set, In the division process, randomization can ensure a consistent data proportions and distribution.

Table 10 illustrates the findings of 6 hybrid deep learning models and their corresponding performance evaluation indexes on the test, including sensitivity (Se), positive predictive value (PPV), accuracy (Acc), negative predictive value (NPV), specificity (Sp), test loss, PPV, NPV, F1-score, MSC, RMSE, MAE, Cohen-kappa score. The classification accuracy of 1D-CNN-GRU, 1D-CNN-LSTM, OGRU-LSTM, GRU-LSTM, CNN-BiLSTM, and RNN-LSTM for 30 % test-samples is 98.90, 95.32, 92.63, 89.27, 87.01 and 84.63 % correspondingly. The 1D-CNN-LSTM and 1D-CNN-GRU classification models clearly show extremely good results. The stacked 1D-CNN-GRU model, in particular, achieved well in terms of assessment indicators: Sensitivity, Specificity, PPV, NPV, F1-score, MSC, RMSE, MAE, and Cohen-kappa score respectively, were 95.27, 94.91, 95.27, 94.60, 0.969, 0.0230, 0.1517, 0.0693, 0.941, with all assessment indicators over 94 %. The stacked 1D-CNN-LSTM classification model has Sensitivity and PPV values of 94.90 and 95.27 %, with a better result, but in terms of Specificity, it was sacrificed, and its NPV and accuracy were 94.40 and 95.32 % respectively. The RNN-LSTM and OGRU-LSTM models have sensitivity results of 87.77 and 90.60 %, correspondingly. To perform data visualization and analysis, **Figure 10** shows a 2-dimensional feature space of the spectral bands of Raman signals. In the developed feature space, the Multi-dimensional Scaling (MDS) graph clearly shows the resemblance between BC and healthy samples. Where the red cross denotes the BC class and the green denotes the control class. The distance among each sample denotes their resemblance in this feature-space, as the distance is nearer, the resemblance is significantly higher and as the distance is higher, the resemblance is lower. As the MDS graph shows, the majority of the samples, which come from various classes can be efficiently differentiated in this feature-space. However, due to a minimum number of overlapping data samples, it is not possible to classify all of the spectral samples reliably using only the limited spectral information from the Raman spectral bands. Since such samples are very close, indicates the relation among them in this feature space is identical. The rest of the Raman spectral bands are merged and the multi-dimensional feature-space is produced, as in each classification part, to classify the samples' overlapping. The receiver operating characteristic (ROC) curve was created to further examine the diagnostic performance of the 6 ML models, as in **Figure 11**. The area under the curve (AUC) indicates the diagnosis performance of the model. The higher the AUC, the better the model's diagnostic performance. The values of AUC for 1D-CNN-GRU, 1D-CNN-LSTM, OGRU-LSTM, GRU-LSTM, CNN-BiLSTM, and RNN-LSTM are 0.9468, 0.9263, 0.8407, 0.8881, 0.9199 and 0.8859, respectively.

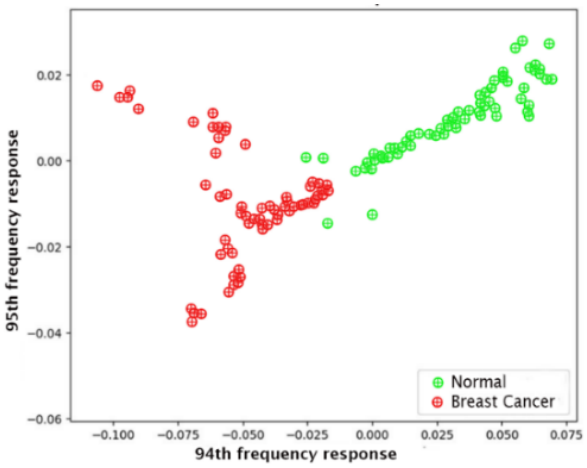


Figure 10 The scatter-plots for BC class and normal class are based on spectral characteristics on the 1D-CNN-GRU classification model.

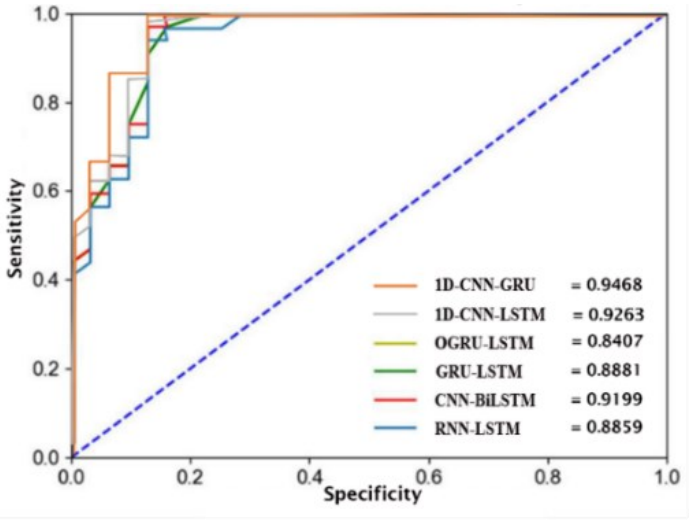


Figure 11 ROC curves of BC against different ML models.

Table 10 Performance of breast cancer classification on different models.

Parameters of evaluation	RNN-LSTM	CNN-BiLSTM	GRU-LSTM	OGRU-LSTM	1D-CNN-LSTM	1D-CNN-GRU
Test Loss	0.435589	0.535684	0.469712	0.486341	0.333505	0.102776
Accuracy (%)	84.63	87.01	89.27	92.63	95.32	98.90
Sensitivity	87.77	89.70	88.90	90.60	94.90	95.27
Specificity	86.10	86.80	85.30	87.00	90.90	94.91
PPV	84.10	91.40	94.20	86.50	94.70	95.27
NPV	94.20	94.30	89.10	91.70	94.40	94.60
F1-score	0.597	0.753	0.767	0.846	0.941	0.969
MSE	0.0293	0.0287	0.0285	0.0292	0.0282	0.0230
RMSE	0.1699	0.1696	0.1670	0.1709	0.1681	0.1517
MAE	0.0893	0.0862	0.0851	0.0863	0.0843	0.0693
Cohen-kappa score	0.603	0.697	0.813	0.884	0.903	0.941

According to literature research, the traditional DL and ML methods used for cancer identification mainly include LDA (linear discriminant analysis), PCA (Principal Component Analysis), SVM (Support Vector Machine), random forest (RF), QDA (Quadratic Discriminant Analysis), PLS (Partial least squares regression), ANN (Artificial neural network) and CNN (Convolution Neural Network). Among them, PCA is applied for dimensionality reduction, while LDA, QDA, SVM, PLS, RF, and CNN are used for spectral signal classification. In some cases, RNN is also used for spectral classification. As shown in **Table 1**, among the conventional ML methods, the deep learning classifier achieves the best evaluation results with accuracy ranging from 94 - 99.5 % for CNN, which is higher than conventional machine learning approaches, and for ANN the accuracy is 86 %. Table 1 demonstrates that the ensemble classifiers with 216 spectra CNN-SVM has better spectral prediction ability than other classification model discussed. Then, the accuracy of GA-PLS-LDA with 2,126 spectra is 98 %. For the linear indistinguishable Raman spectral data used in the experiments, the selection of the kernel function and reasonable parameter adjustment is crucial. Additionally, the CNN classifier performed on 15,354 spectra is evaluated with 94 - 99 % accuracy. For the Raman spectral analysis using 12,000 spectra, VGG-16 reached an accuracy of 92.6 %. In the case of SVM, for 4,860 spectra around 92.89 % accuracy is obtained on the other hand SVM has accuracy ranging from 72 to 90 %. ANN has a significantly low accuracy rate of 86 % against 928 spectra. It is analyzed from the literature research that both the classification models and the number of spectra influence the prediction results. The better the prediction results the number of spectral signatures to be trained by the model should be higher. Also, it is evident that the deep learning ensemble model has comparatively higher classification results. Finally, the experimental results of the proposed system after evaluating with other baseline hybrid deep learning models experimented with 62,199 spectra, 1D-CNN-GRU has an accuracy of 98.90, 0.969 % of F1-score, 0.941 % of Cohen-kappa score, and 0.0230 MSE with a satisfactory result. The capability of a hybrid deep learning model 1D-CNN-GRU on high dimensional spectral data is potent, comparatively better performances are achieved from other tested hybrid deep learning models. Based on the comparative evaluation results, that not all deep learning models are better than the conventional machine learning models for Raman spectral signal classification, but choosing a suitable combination of deep learning networks called hybrid deep learning models, will greatly boost the classification performance of the models. Compared with deep learning methods, hybrid deep learning can exploit the potential features of data as much as possible.

Conclusions

Breast neoplasm is a serious disease that must be treated with care. Early detection of this condition is extremely beneficial in saving millions of lives. The objective of this research work is to see if a Raman spectral feature can be used to assess the probability of being affected by BC disease at an early stage with high sensitivity and specificity by comparing it with the performance of hybrid deep learning models. This research, proposed and implements 6 different hybrid deep learning models with Raman spectral data. The 1D-CNN-GRU hybrid model has proved its ability with maximum efficiency for early diagnosis of BC using blood plasma samples obtained from clinically confirmed BC candidates. When developing the model with the necessary parameter tuning, interfering characteristics that have an impact on this condition were considered. The stacked 1D-CNN-GRU model achieved the highest classification accuracy among other hybrid deep-learning classifiers tested with the same spectral samples. Accuracy, Cohen-kappa score, F1-score, lowermost number of test losses, and MSE are indicating that the model outperforms other hybrid deep learning classifiers with 98.90 % accuracy, 0.969 F1-score, 0.941 Cohen-kappa score, and 0.0230 MSE with a satisfactory result. Five efficient HDL algorithms were proposed along with the 1D-CNN-GRU approach such as 1D-CNN-LSTM, OGRU-LSTM, GRU-LSTM, CNN-BiLSTM, and RNN-LSTM. These are employed to develop and compare the performance of automated BC diagnosis models for the classification of spectral data. The accuracy, specificity, sensitivity, ROC curve NPV, and PPV were employed as assessment criteria to estimate the performance of the classification of these 6 algorithms against these spectral data at the same time. The findings show that the 1D-CNN-LSTM classification model performs well after the 1D-CNN-GRU model in terms of the overall efficiency of 95.32 %.

References

[1] F Bray, J Ferlay, I Soerjomataram, RL Siegel, LA Torre and A Jemal. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Canc. J. Clinicians* 2018; **68**, 394-424.

[2] W Gordenne. Mammography: The gold standard of breast mass screening. *J. Belg. Radiol.* 1990; **73**, 335-8.

[3] HG Welch, LM Schwartz and S Woloshin. Ramifications of screening for breast cancer: 1 in 4 cancers detected by mammography are pseudocancers. *BMJ* 2006; **332**, 727.
[4] GW Auner, SK Koya, C Huang, B Broadbent, M Trexler, Z Auner, A Elias, KC Mehne and MA Brusatori. Applications of Raman spectroscopy in cancer diagnosis. *Canc. Metastasis Rev.* 2018; **37**, 691-717.
[5] R Manoharan, Y Wang and MS Feld. Histochemical analysis of biological tissues using Raman spectroscopy. *Spectrochim. Acta Mol. Biomol. Spectros.* 1996; **52**, 215-49.
[6] J Surmacki, J Musial, R Kordek and H Abramczyk. Raman imaging at biological interfaces: applications in breast cancer diagnosis. *Mol. Canc.* 2013; **12**, 48.
[7] AS Haka, KE Shafer-Peltier, M Fitzmaurice, J Crowe, RR Dasari and MS Feld. Identifying microcalcifications in benign and malignant breast lesions by probing differences in their chemical composition using Raman spectroscopy. *Canc. Res.* 2002; **62**, 5375-80.
[8] VS Renjith and PSH Jose. A noninvasive approach using multi-tier deep learning classifier for the detection and classification of breast neoplasm based on the staging of tumor growth. *In*: Proceedings of the 2020 International Conference on Decision Aid Sciences and Application, Sakheer, Bahrain. 2020, p. 471-8.
[9] LA Austin, S Osseiran and CL Evans. Raman technologies in cancer diagnostics. *Analyst* 2016; **141**, 476-503.
[10] K Kong, C Kendall, N Stone and I Notingher. Raman spectroscopy for medical diagnostics-from *in-vitro* biofluid assays to *in-vivo* cancer detection. *Adv. Drug Deliv. Rev.* 2015; **89**, 121-34.
[11] U Neugebauer, P Rösch and J Popp. Raman spectroscopy towards clinical application: Drug monitoring and pathogen identification. *Int. J. Antimicrob. Agents* 2015; **46**, S35-9.
[12] MM Joseph, N Narayanan, JB Nair, V Karunakaran, AN Ramya, PT Sujai, G Saranya, JS Arya, VM Vijayan and KK Maiti. Exploring the margins of SERS in practical domain: An emerging diagnostic modality for modern biomedical applications. *Biomaterials* 2018; **181**, 140-81.
[13] MJ Winterhalder and A Zumbusch. Beyond the borders-biomedical applications of non-linear raman microscopy. *Adv. Drug Deliv. Rev.* 2015; **89**, 135-44.
[14] G Agarwal, P Pradeep, V Aggarwal, CH Yip and PS Cheung. Spectrum of breast cancer in Asian women. *World J. Surg.* 2007; **31**, 1031-40.
[15] JRF Caldeira, ÉC Prando, FC Quevedo, FAM Neto, CA Rainho and SR Rogatto. CDH1 promoter hypermethylation and e-cadherin protein expression in infiltrating breast cancer. *BMC Canc.* 2006; **6**, 48.
[16] P Pradhan, S Guo, O Ryabchykov, J Popp and TW Bocklitz. Deep learning a boon for biophotonics? *J. Biophotonics* 2020; **13**, e201960186.
[17] MA Mohammed, B Al-Khateeb, AN Rashid, DA Ibrahim, MKA Ghani and SA Mostafa. Neural network and multi-fractal dimension features for breast cancer classification from ultrasound images. *Comput. Electr. Eng.* 2018; **70**, 871-82.
[18] RDH Devi and P Deepika. Performance comparison of various clustering techniques for diagnosis of breast cancer. *In*: Proceedings of the 2015 IEEE International Conference on Computational Intelligence and Computing Research, Madurai, India. 2015.
[19] PJ Sudharshan, C Petitjean, F Spanhol, LE Oliveira, L Heutte and P Honeine. Multiple instances learning for histopathological breast cancer image classification. *Expert Syst. Appl.* 2019; **117**, 103-11.
[20] ZC Lipton, J Berkowitz and C Elkan. A critical review of recurrent neural networks for sequence learning. *arXiv* 2015; **1**, 1-38.
[21] J Dong, M Hong, Y Xu and X Zheng. A practical convolutional neural network model for discriminating Raman spectra of human and animal blood. *J. Chemometr.* 2019; **33**, e3184
[22] W Lee, ATM Lenferink, C Otto and HL Offerhaus. Classifying Raman spectra of extracellular vesicles based on convolutional neural networks for prostate cancer detection. *J. Raman Spectros.* 2020; **51**, 293-300.
[23] MK Maruthamuthu, AH Raffiee, DMD Oliveira, AM Ardekani, and MS Verma. Raman spectra-based deep learning: A tool to identify microbial contamination. *MicrobiologyOpen* 2020; **9**, e1122.
[24] S Boonsit, P Kalasuwan, PV Dommelen and C Daengngam. Rapid material identification via low-resolution Raman spectroscopy and deep convolutional neural network. *J. Phys. Conf.* 2021; **1719**, 012081.
[25] N Cheng, J Fu, D Chen, S Chen and H Wang. An antibody-free liver cancer screening approach based on nanoplasmonics biosensing chips via spectrum-based deep learning. *NanoImpact* 2021; **21**, 100296.

[26] J Houston, FG Glavin and MG Madden. Robust classification of high-dimensional spectroscopy data using deep learning and data synthesis. *J. Chem. Inform. Model.* 2020; **60**, 1936-54.
[27] CS Ho, N Jean, CA Hogan, L Blackmon, SS Jeffrey, M Holodniy, N Banaei, AAE Saleh, S Ermon and J Dionne. Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning. *Nat. Comm.* 2019; **10**, 4927.
[28] J Ding, M Yu, L Zhu, T Zhang, J Xia and G Sun. Diverse spectral band-based deep residual network for tongue squamous cell carcinoma classification using fiber optic Raman spectroscopy. *Photodiagnosis Photodynamic Ther.* 2020; **32**, 102048.
[29] H Chen, C Chen, H Wang, C Chen, Z Guo, D Tong, H Li, H Li, R Si, H Lai and X Lv. Serum Raman spectroscopy combined with a multi-feature fusion convolutional neural network diagnosing thyroid dysfunction. *Optik* 2020; **216**, 164961.
[30] CM Krishna, GD Sockalingum, J Kurien, L Rao, L Venteo, M Pluot, M Manfait and VB Kartha. Micro-Raman spectroscopy for optical pathology of oral squamous cell carcinoma. *Appl. Spectros.* 2004; **58**, 1128-35.
[31] MS Bergholt, W Zheng and Z Huang. Characterizing variability in *in vivo* Raman spectroscopic properties of different anatomical sites of normal tissue in the oral cavity. *J. Raman Spectros.* 2012; **43**, 255-62.
[32] SP Singh, A Deshmukh and P Chaturvedi. *In vivo* Raman spectroscopic identification of premalignant lesions in oral buccal mucosa. *J. Biomed. Optic.* 2012; **17**, 105002.
[33] LP Choo-Smith, HGM Edwards, HP Endtz, JM Kros, F Heule, H Barr, JS Robinson, HA Bruining and GJ Puppels. Medical applications of Raman spectroscopy: From proof of principle to clinical implementation. *Biopolymers* 2002; **67**, 1-9.
[34] C Kallaway, LM Almond, H Barr, J Wood, J Hutchings, C Kendall and N Stone. Advances in the clinical application of Raman spectroscopy for cancer diagnostics. *Photodiagnosis Photodynamic Ther.* 2013; **10**, 207-19.
[35] P Zuvela, K Lin, C Shu, W Zheng, C Lim and Z Huang. Fiber-optic Raman spectroscopy with nature-inspired genetic algorithms enhances real-time *in vivo* detection and diagnosis of nasopharyngeal carcinoma. *Anal. Chem.* 2019; **91**, 8101-8.
[36] M Yu, H Yan, J Xia, L Zhu, T Zhang, Z Zhu, X Lou, G Sun and M Dong. Deep convolutional neural networks for tongue squamous cell carcinoma classification using Raman spectroscopy. *Photodiagnosis and Photodynamic Ther.* 2019; **26**, 430-5.
[37] J Xia, L Zhu, M Yu, T Zhang, Z Zhu, X Lou, G Sun and M Dong. Analysis and classification of oral tongue squamous cell carcinoma based on Raman spectroscopy and convolutional neural networks. *J. Mod. Optic.* 2020; **67**, 481-9.
[38] C Shu, H Yan, W Zheng, K Lin, A James, S Selvarajan, C Lim and Z Huang. Deep learning-guided fiberoptic raman spectroscopy enables real-time *in vivo* diagnosis and assessment of nasopharyngeal carcinoma and post-treatment efficacy during endoscopy. *Anal. Chem.* 2021; **93**, 10898-906.
[39] T Sciortino, R Secoli, E D'Amico, S Moccia, MC Nibali, L Gay, M Rossi, N Pecco, A Castellano, ED Momi, B Fernandes, M Riva and L Bello. Raman spectroscopy and machine learning for IDH genotyping of unprocessed glioma biopsies. *Cancers* 2021; **13**, 4196.
[40] M Riva, T Sciortino, R Secoli, E D'Amico, S Moccia, B Fernandes, MC Nibali, L Gay, M Rossi, ED Momi and L Bello. Glioma biopsies classification using raman spectroscopy and machine learning models on fresh tissue samples. *Cancers* 2021; **13**, 1073.
[41] K Mehta, A Atak, A Sahu and S Srivastava. An early investigative serum Raman spectroscopy study of meningioma. *Analyst* 2018; **143**, 1916-23.
[42] K Liua, B Liua, Y Zhang, Q Wu, M Zhong, L Shang, Y Wang, P Liang, W Wang, Q Zhao and B Li. Building an ensemble learning model for gastric cancer cell line classification via rapid raman spectroscopy. *Comput. Struct. Biotechnol. J.* 2023; **21**, 802-11.
[43] MS Silva-López, CAI Hernández, HRN Contreras, ÁGR Vázquez, A Ortiz-Dosal and ES Kolosovas-Machuca. Raman spectroscopy of individual cervical exfoliated cells in premalignant and malignant lesions. *Appl. Sci.* 2022; **12**, 2419.
[44] Z Liu, T Li, Z Wang, J Liu, S Huang, BH Min, JY An, KM Kim, S Kim, Y Chen, H Liu, Y Kim, DTW Wong, TJ Huang and X Ya-Hong. Gold nanopyramid arrays for non-invasive surface-enhanced raman spectroscopy-based gastric cancer detection via sEVs. *ACS Appl. Nano Mater.* 2022; **5**, 12506-17.
[45] L Huang, H Sun, L Sun, K Shi, Y Chen, X Ren, Y Ge, D Jiang, X Liu, W Knoll, Q Zhang and Y Wang. Rapid, label-free histopathological diagnosis of liver cancer based on Raman spectroscopy and deep learning. *Nat. Comm.* 2023; **14**, 48.

[46] M Jeng, LCT Sharma, S Huang, L Chang, S Wu and L Chow. Raman spectroscopy analysis for optical diagnosis of oral cancer detection. *J. Clin. Med.* 2019; **8**, 1313.
[47] C He, X Wu, J Zhou, Y Chen and J Ye. Raman optical identification of renal cell carcinoma via machine learning. *Spectrochim. Acta Mol. Biomol. Spectros.* 2021; **252**, 119520.
[48] C Chen, W Wu, C Chen, F Chen, X Dong, M Ma, Z Yan, X Lv, Y Ma and M Zhu. Rapid diagnosis of lung cancer and glioma based on serum Raman spectroscopy combined with deep learning. *J. Raman Spectros.* 2021; **52**, 1798-809.
[49] D Bury, G Faust, M Paraskevaidi, KM Ashton, TP Dawson and FL Martin. Phenotyping metastatic brain tumors applying spectrochemical analyses: Segregation of different cancer types. *Anal. Lett.* 2019; **52**, 575-87.
[50] K Aubertin, VQ Trinh, M Jermyn, P Baksic, AA Grosset, J Desroches, K St-Arnaud, M Birlea, MC Vladoiu, M Latour, R Albadine, F Saad, F Leblond and D Trudel. Mesoscopic characterization of prostate cancer using Raman spectroscopy: Potential for diagnostics and therapeutics. *BJU Int.* 2018; **122**, 326-36.
[51] H Yan, M Yu, J Xia, L Zhu, T Zhang, Z Zhu and G Sun. Diverse region-based CNN for tongue squamous cell carcinoma classification with Raman spectroscopy. *IEEE Access* 2020; **8**, 127313-28.
[52] X Wu, S Li, Q Xu, X Yan, Q Fu, X Fu, X Fang and Y Zhang. Rapid and accurate identification of colon cancer by Raman spectroscopy coupled with convolutional neural networks. *Jpn. J. Appl. Phys.* 2021; **60**, 067001.
[53] IP Santos, PV Doorn, PJ Caspers, TCB Schut, EM Barroso, TE Nijsten, VN Hegt, S Koljenovi´c and GJ Puppels. Improving clinical diagnosis of early-stage cutaneous melanoma based on Raman spectroscopy. *Br. J. Canc.* 2018; **119**, 1339.
[54] JD Gelder, KD Gussem, P Vandenabeele and L Moens. Reference database of Raman spectra of biological molecules. *J. Raman Spectros.* 2007; **38**, 1133-47.
[55] AD Meade, FM Lyng, P Knief and HJ Byrne. Growth substrate induced functional changes elucidated by FTIR and Raman spectroscopy in *in-vitro* cultured human keratinocytes. *Anal. Bioanalytical Chem.* 2007; **387**, 1717-28.
[56] I Notingher. Raman spectroscopy cell-based biosensors. *Sensors* 2007; **7**, 1343-58.
[57] P Jess, V Garcés-Chávez, D Smith, M Mazilu, L Paterson, A Riches, C Herrington, W Sibbett and K Dholakia. Dual beam fibre trap for Raman microspectroscopy of single cells. *Optic. Express* 2006; **14**, 5779-91.
[58] Z Movasaghi, S Rehman and IU Rehman. Raman spectroscopy of biological tissues. *Appl. Spectros. Rev.* 2007; **42**, 493-541.
[59] H Nawaz, F Bonnier, P Knief, O Howe, FM Lyng, AD Meade and HJ Byrne. Evaluation of the potential of Raman microspectroscopy for prediction of chemotherapeutic response to cisplatin in lung adenocarcinoma. *Analyst* 2010; **135**, 3070-6.
[60] MH Mozaffari and T Li-Lin. Overfitting one-dimensional convolutional neural networks for Raman spectra identification. *Spectrochim. Acta Mol. Biomol. Spectros.* 2022; **272**, 120961.
[61] A Shrivastava, T Pfister, O Tuzel, J Susskind, W Wang and R Webb. Learning from simulated and unsupervised images through adversarial training. *In*: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii. 2017.
[62] I Goodfellow, Y Bengio and A Courville. *Deep learning*. MIT press: Cambridge, Massachusetts, 2016.
[63] J Chen, X Hu and Y Liu. Improved gated recurrent units for time series prediction. *In*: Proceedings of the 2018 IEEE International Conference on Big Data, Seattle, Washington. 2018
[64] Y Liu, S Chen and J Li. *Deep learning for time series modelling*. arXiv, Illinois, 2015.
[65] MF Karim, MJ Hossain and A Svyatkovskiy. Multivariate time series forecasting using 1D convolutional neural networks. *In*: Proceedings of the 2019 International Conference on Computer and Communication Engineering, Pune, India. 2019.
[66] DM Durairaj and BHK Mohan. A convolutional neural network-based approach to financial time series prediction. *Neural Comput. Appl.* 2022; **34**, 13319-37.
[67] X Wang, J Xu, W Shi and J Liu. OGRU: An optimized gated recurrent unit neural network. *J. Phys. Conf.* 2019; **1325**, 012089.
[68] J Du, Y Cheng, Q Zhou, J Zhang, X Zhang and G Li. Power load forecasting using BiLSTM-attention. *IOP Conf. Ser. Earth Environ. Sci.* 2020; **440**, 032115.
[69] D Gjylapi and E Proko. Recurrent neural networks in time series prediction. *In*: Proceedings of the New Research and Advances on Computer Science and Information Technology NRACSIT-2017, Vlora, Albania. 2017.

[70] F Chollet. Xception: Deep Learning with Depthwise Separable Convolutions. *In*: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017, p. 1800-7.
[71] M Abadi, P Barham, J Chen, Z Chen, A Davis, J Dean, M Devin, S Ghemawat, G Irving, M Isard, M Kudlur, J Levenberg, R Monga, S Moore, DG Murray, B Steiner, P Tucker, V Vasudevan, P Warden, M Wicke, Y Yu and X Zheng. TensorFlow: A system for large-scale machine learning. *arXiv* 2016; **1605**, 08695.
[72] DO Oyewola, EG Dada, S Misra and R Damaševicius. A novel data augmentation convolutional neural network for detecting malaria parasite in blood smear images. *Appl. Artif. Intell.* 2022; **36**, 2033473.
[73] DO Oyewola, EG Dada and O Ezekiel. Predicting nigerian stock returns using technical analysis and machine learning. *Eur. J. Electr. Eng. Comput. Sci.* 2019; **3**, 1-8.
[74] D Yao, B Li, H Liu, J Yang and L Jia. Remaining useful life prediction of roller bearings based on improved 1D-CNN and simple recurrent unit. *Measurement* 2021; **175**, 109166.
[75] M Wei, H Gu, M Ye, Q Wang, X Xu and C Wu. Remaining useful life prediction of lithium-ion batteries based on monte carlo dropout and gated recurrent unit. *Energ. Rep.* 2021; **7**, 2862-71.
[76] J Zhang, F Wei, F Feng and C Wang. Spatial-spectral feature refinement for hyperspectral image classification based on attention-dense 3D-2D-CNN. *Sensors* 2020; **20**, 5191.
[77] Y Zhang, M Liu, L Kong, T Peng, D Xie, L Zhang, L Tian and X Zou. Temporal characteristics of stress signals using GRU algorithm for heavy metal detection in rice based on sentinel-2 images. *Int. J. Environ. Res. Publ. Health* 2022; **19**, 2567.
[78] QJ Lv, HY Chen, WB Zhong, YY Wang, JY Song, SD Guo, LX Qi and CY Chen. A multi-task group Bi-LSTM networks application on electrocardiogram classification. *IEEE J. Translational Eng. Health Med.* 2019; **8**, 1900111.