Foreseeing Survival Through 'Fuzzy Intelligence': A Cognitively-Inspired Incremental Learning Based *de novo* Model for Breast Cancer Prognosis by Multi-Omics Data Fusion

Aviral Chharia, Neeraj Kumar

Thapar Institute of Engineering and Technology, Patiala, PB 147004, India
{achharia_be18, neeraj.kumar}@thapar.edu

Abstract. High-precision breast cancer prognosis is crucial for early disease identification, avoiding hazardous side effects of unnecessary therapies, and decreasing mortality rates through personalized and tailored treatment regimens. However, designing a prognosis model continues to be challenging, given the intricate relationship between distinct genetic attributes, varied clinical results of drug therapies, the noisy nature of gene expressions, and the high-class imbalance seen in multimodal cancer data. Furthermore, because labeled omics data collection is costly and requires highly trained experts, the data available is very limited. This makes the design of the conventional machine and deep learning models incredibly challenging, as they require large quantities of data for learning the underlying intricate patterns and would otherwise overfit, decreasing model precision. Moreover, all present models suffer from a 'closed world assumption.' These models, once trained, cannot be updated in real-time (when more omics data is available in the future) without a complete re-training. The present study is the first to introduce the 'Fuzzy' way towards Breast cancer prognosis, framing the task as an incremental learning problem. The proposed approach allows the model to continually update its learned feature space on a non-stationary multimodal data stream, emulating the human brain's remarkable quality to learn over time. We demonstrate the model's ability to learn complex relationships between different multimodal attributes, training on severely imbalanced and limited data by mapping it to a high-dimensional 'fused' feature space. The proposed model surpasses state-of-the-art machine learning models significantly. These results suggest that prediction through 'fuzzy intelligence' is a promising approach towards breast cancer prognosis.

Jul 1, 2021: This author's *preprint* reports new research that has not yet undergone peer review and should not be used to guide clinical practice. It is released to ensure timely dissemination of scholarly work and may differ from final version, which will undergo review and editorial processing.
Sept 25, 2021: Final peer-reviewed and accepted version is now available online: https://doi.org/10.1007/978-3-030-87602-9_22. It was presented at the 4th Predictive Intelligence in Medicine Workshop, held in Conjunction with MICCAI, Strasbourg, France, 2021. Please refer to and cite the Springer version in your research.
© 2021 Springer Nature Switzerland AG. Permission from Springer must be obtained for all uses, in any current or future media.

1 Introduction

According to the American Society of Oncology, Breast Cancer is the most frequently occurring cancer amongst women. In 2020, breast cancer affected 276,480 women in the United States alone. The development of a Breast cancer prognosis model can help oncologists offer personalized treatment plans, especially in the case of the more aggressive invasive breast cancer, which spreads within the body, reducing 5-year survival rates to as low as 27%. Moreover, an effective prognosis can help to increase life expectancy, especially if the patient is a short-term survivor. Breast cancer prognosis is challenging due to its underlying heterogeneous nature and high complexity. Patients in the same stage and with similar clinical characteristics often undergo different therapies, with disparate responses in most cases affecting overall survival. This makes the design of a correlative prognosis model challenging and, therefore, of practical interest to oncologists. Recent epidemiological and linkage research has established that mutations in genes and a lack of 'alleles' in the BRCA1 locus increase susceptibility to breast cancer. Moreover, gene expression and DNA copy number alteration (CNA) data are quite noisy and large (order of 2×10^4 features). However, the number of training samples is very less, making patients' characterization as long-term ($>$ 5-year) and short-term survivors ($<$ 5-year) a challenging task.

Related Works. Several studies have focused on predicting survival rates in patients diagnosed with breast cancer, but have certain limitations. [1] was the first to propose a prognosis model, but used only gene expression profiles. Recent developments in high-throughput microarrays and gene expression technologies have shown that gene signatures are not the only contributing factor in breast cancer. Assuming different genes of a particular patient may have significant relations amongst themselves, [2] used a support vector machine (SVM), while [3] used Random Forest coupled with efficient feature selection for high accuracy prognosis. The deep belief network proposed by [4] combined two independent microarray data, i.e., gene expression and clinical, while using principal component analysis (PCA) for dimensionality reduction. The major limitation of these models was that they assumed that different modalities have the same feature representations. Recently, deep learning-based supervised feature extraction has gained immense attention [5]. Recent work by [6] used a score-level fusion of coefficients for integrating multi-modal data. However, the major disadvantage of the model was that these coefficients need to be manually determined, which is an iterative and challenging task in itself. Work by [7] combined CNN-based stacked feature extraction with various machine learning models. The major limitation of this work was that the proposed method used models that are quite data-intensive and cannot work on severely imbalanced datasets. Also, the study lacks cross-validation on larger datasets required for deep learning models. Furthermore, since these models are trained on limited data, they may be prone to overfitting.

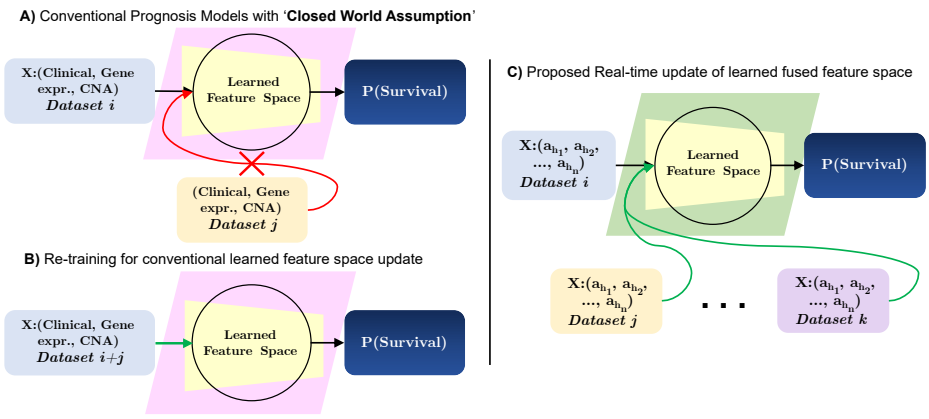


Fig. 1. Illustration of the **A.** Conventional prognosis models with 'closed world assumption' **B.** The required re-training for conventional leaned feature space update **C.** Proposed real-time update $\mathbf{X} : (a_{h_1}, a_{h_2}, ..., a_{h_n}) \rightarrow \mathbf{i}, \mathbf{j}, \mathbf{k}$ of learned fused feature space

Contributions. We summarize our contributions as follows:

1. The present work is the *first* to formulate Cancer Prognosis as an incremental learning problem combining multimodal omics data and proposes a *de novo* cognitively-inspired 'fuzzy' network for breast cancer prognosis. The proposed model outperforms other state-of-the-art machine learning models by a significant margin. Furthermore, as compared to prior techniques, real-time architecture update eliminates the need for model re-training if new labeled data is available in the future.
2. The model's other novelty lies in its capacity to attain high classification accuracy despite being trained on limited data samples, as proven experimentally, in contrast to previously proposed models that need large amounts of multi-omics data difficult and costly to obtain.
3. Another unique feature of the model lies in its robustness to high-class imbalance, commonly seen in real-world multi-omics datasets, as demonstrated in the experiments performed. On the contrary, as the class imbalance in the dataset increases, most machine learning and deep learning models exhibit a significant reduction in classification accuracy.

2 Proposed Methodology

Figure 2 illustrates the architecture of the proposed breast cancer prognosis model. This section discusses in detail the proposed methodology.

Weighted k-NN and mRMR Feature Selection. *Firstly*, the Weighted nearest neighbor algorithm [8] is used for estimating the missing gene expression and CNA profiles in the dataset. *Secondly*, feature selection using the mRMR

(Minimum redundancy maximum relevance) algorithm [9,10], which reduces dimensionality without significant data loss, was used to escape the curse of dimensionality. Here, gene expression profiles were reduced from 24368 $\rightarrow$ 400, CNA from 26298 $\rightarrow$ 200, and clinical from 27 $\rightarrow$ 25 [6]. Further, the gene expression features are normalized and discretized into three categories: under-expression (-1), over-expression (1), and baseline (0), i.e., $\in \{-1, 0, 1\}$ [9]. The CNA features are discretized $\in \{-2, -1, 0, 1, 2\}$ and clinical data is normalized $\in [0, 1]$ using min-max normalization.

Stacked CNN for Multimodal Feature Extraction and Fusion. Since different data modalities may have different feature representations, the direct feature fusion of multi-sourced data to a deep neural net may not be ideal. Therefore, separate CNN models [7] are used for each: clinical, gene expression, and CNA. Each of the CNNs is trained on a single METABRIC modality with **AUC** value as the evaluation metric. Binary-cross entropy is used as the loss function $\mathcal{L}$ with **L2** regularization and learning rate 10^{-3} for 8 mini-batches and 20 training epochs as,

$$\mathcal{L}(y_t, \hat{y}_t) = -\frac{1}{\mathcal{N}} \sum_{i=0}^{\mathcal{N}} \left[y_t(i) \log \hat{y}_t(i) - (1 - y_t(i)) \log(1 - \hat{y}_t(i)) \right] + \frac{1}{2} \lambda \sum_{k=1}^{\mathbf{K}} \sum_{j=1}^{\mathbf{n_k}} \sum_{i=1}^{\mathbf{m_k}} w_{ij}^{k^2}$$

where $\mathcal{N}$ is the batch size, $\mathbf{K}$ is the number of weight matrices in the CNN, $\mathbf{W}^k = (w_{ij}^k)_{(\mathbf{m_k} \times \mathbf{n_k})}$ is the k_{th} weight matrix, y_t and $\hat{y}_t$ are the actual and predicted labels. A feature map, i.e., element-wise $\odot$ followed by addition between the filter matrix and corresponding values of the input matrix, is produced. Glorot normal initializer [12] is used for filter matrix initialization. It selects random numbers with mean = 0 and standard deviation in, $\left[-\sqrt{\frac{2}{n_i+n_o}}, \sqrt{\frac{2}{n_i+n_o}}\right]$, where n_i and n_o represents number of input and output units for selected layer, respectively. For the Convolution layer, 4 filters were used with size 15, and the stride size was 2. The introduced hidden layer had 150 units. The obtained hidden feature vectors from each trained CNN are fused to get stacked features.

Mapping *n*-dimensional 'Fused' Feature Space. The obtained stacked feature vector $\{a_h, C_i\}$ for each patient is passed to the input nodes $\{a_1, ..., a_h\}$ of the fuzzy classifier [13] after normalization $\in [0.01, 0.99]$. The fuzzy classifier works by creating hyperboxes $\mathcal{H}$ [14], which is a geometrical shape defined in the n-dimensional feature space. Parameters $V_j = (v_{j1}, v_{j2}, ..., v_{jn})$ and $W_j = (w_{j1}, w_{j2}, ..., w_{jn})$ are used to define the min and max-coordinates of $\mathcal{H}$, while 'θ' i.e., the hyperbox expansion coefficient $\in (0, 1)$ represents its size and 'γ' represents the fuzziness control parameter.

Incremental Learning Based Model Training. Figure 1 illustrates the incremental learning approach, on which the proposed model is based. For each input feature vector, Classifying Neurons (CLN) performs the classification

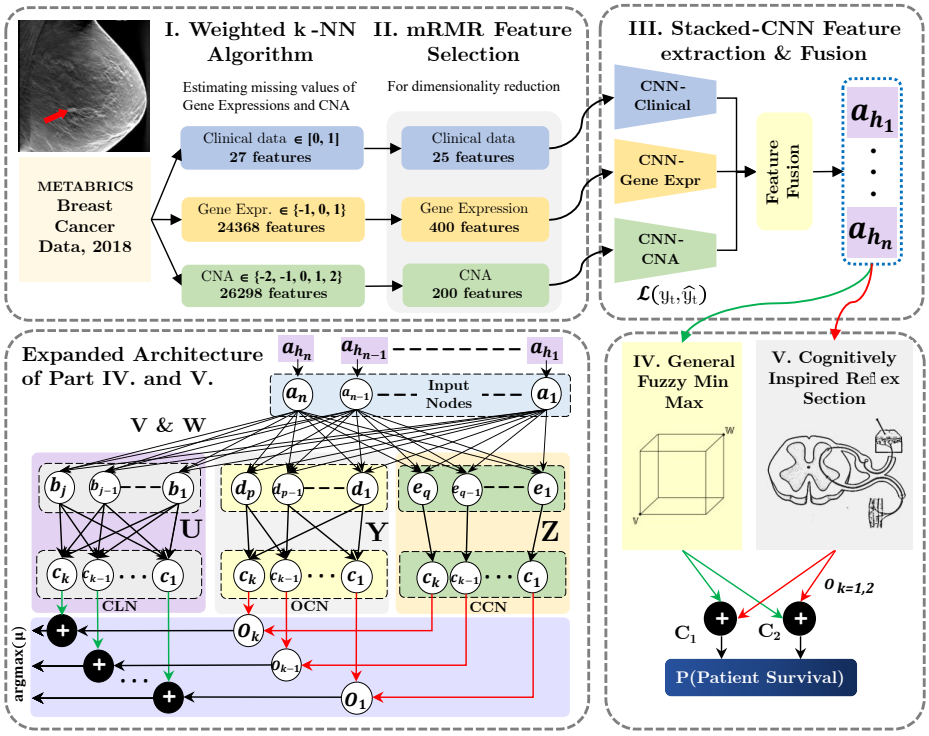


Fig. 2. Architecture of the proposed breast cancer prognosis model based on 'Fuzzy' Incremental Learning for classifying patients as long and short-term survivors on multimodal omics (i.e., gene expressions, copy number alteration, and clinical) data

of learned data using min-max hyperboxes. In CLNs, neuron b_j represents hyperbox fuzzy set $B_j = A_h, V_j, W_j, f(A_h, V_j, W_j) \forall (A_h \in I^n)$. To compute class memberships, the activation function by [15] is used to assign a membership value $= 1$ when the test sample falls within $\mathcal{H}$. In other cases, when the test sample lies outside $\mathcal{H}$, the membership value is calculated on the basis of its distance from the extreme co-ordinates of $\mathcal{H}$. The classifying neuron activation function, i.e., b_j is defined as,

$$b_j(a_h, V_j, W_j) = \min_{i=1..n} \Big(\min \Big[\big(1 - f\big(a_{hi} - W_{ji}, \gamma\big)\big), \big(1 - f\big(V_{ji} - a_{hi}, \gamma\big)\big) \Big] \Big)$$

where, $f(x, \gamma) = 0$ if $x\gamma < 0$; 1 if $x\gamma > 1$ and equal to $x\gamma$ if it lies $\in [0, 1]$. In the middle layer, the input nodes and the hyperbox ($\mathcal{H}$) nodes are connected. These connections represents $\mathbf{V}$ and $\mathbf{W}$ of the n-dimensional hyperbox fuzzy set [16]. During training, the neurons in the middle layer are created dynamically and connection between the $\mathcal{H}$ node b_j to a class node C_j, is represented by matrix $\mathbf{U}$, where, $u_{ij} = 1$ if $b_j \in C_j$ else $u_{ij} = 0$. This real-time architecture update allows for the mid-way introduction of new labeled data. More labeled multimodal omics

data, *if available*, in the future may be directly passed through the trained model to update its learned fused feature space, thus enabling Incremental learning.

Whenever a training sample is encountered that doesn't belong to a class the model has learned previously, a $\mathcal{H}$ node is created. During training, the model tries to accommodate subsequent samples $\{a_h, C_i\}$ in the previously made $\mathcal{H}$ belonging to the same class using conditions below [13]. If expansion of any of the existing hyperboxes belonging to that class is not feasible, a new $\mathcal{H}$ is added; i.e., for a new training sample $\{a_h, C_i\}$, a hyperbox $\{b_j, C_j\}$ is found such that $C_j = C_i$ or $C_j = C_0$ which has the highest membership value and satisfying-

1. $\theta_{\max} \geq \frac{1}{n}\sum_{i=1}^{n}(\max(w_{ji}, a_{hi}) - \min(v_{ji}, a_{hi}))$
2. b_j is not associated with any OCN/ CCN
3. if $C_i = C_0$ or $C_j = C_0$ then $\mu_j > 0$, where μ_j is membership with b_j,

Adjust min-max coordinates of b_j, as,
$V_{ji}^{new} = \min(V_{ji}^{old}, a_{hi})$,
$W_{ji}^{new} = \max(W_{ji}^{old}, a_{hi})$, where $i = 1, 2, ..., n$
and if $C_j = C_0$ and $C_i \neq C_0$ then $C_j = C_i$

If no suitable b_j is present, then a novel hyperbox $\mathcal{H}$ for class C_i is created with $V_j = W_j = a_h$; i.e., a point hyperbox.

Cognitively-Inspired Reflex Section. The Reflex section is cognitively-inspired [16] from the human brain and handles cases of $\mathcal{H}$ overlap and containment, which may arise due to visual feature overlap in the high-dimensional feature space. The Overlap Compensation Neuron **(OCN)** becomes active if the test data lies within the overlap space and generates two compensation outputs, one each for the two overlapping classes, with activation function-

$$d_{j_p} = T(b_j(a_h, V_j, W_j) - 1) \times \left(-1 + \frac{1}{n}\sum_{i=1}^{n} \max\left(\frac{a_{hi}}{w_{p_{ji}}}, \frac{v_{p_{ji}}}{a_{hi}}\right)\right)$$

where, $T(x) = 0$ if $x < 0$ and 1 if $x \geq 0$. The Containment Compensation Neuron **(CCN)** overcomes $\mathcal{H}$ containment cases and has activation function-

$$e_j = -1 \times T(b_j(a_h, V, W) - 1)$$

The final class for each sample is computed as the **argmax** of the sum of the membership and compensation values.

3 Experiments and Results

Dataset and Evaluation Metrics. The Molecular Taxonomy of Breast Cancer International Consortium [17] data contains the clinical data, gene expression profiles, and CNA profiles of 1980 breast cancer patients in the METABRIC trial. The pre-processed METABRIC data at (https://github.com/USTC-HIlab/MDNNMD) is used in this study for validating the proposed model. In datasets that

are skewed towards a particular class, accuracy alone may be misleading when assessing model performance. Therefore, on imbalanced datasets, precision along with accuracy is considered as a better evaluation metric [18].

Exp-1. Model Performance Comparison on Limited Data Subset Configuration. In the first experiment, the model is trained on limited data subset configurations, i.e., $n = 200, 300, 400, 500$ (with no class-imbalance), where n is the number of samples. The obtained results are compared with state-of-the-art machine learning models trained on the same number of data samples (see Table 1). Here, confusion-matrix-based evaluation metrics were used for performance comparison. Further, it is to be noted that the same feature selection and pre-processing techniques (Weighted k-NN algorithm and mRMR Feature selection) are used while implementing the machine learning models for a fair comparison of results. The proposed model outperforms all other models with a significant margin in all data subset configurations, i.e., for $n = 200, 300, 400, 500$. On $n = 300$ samples data, the proposed model surpasses the best-performed machine learning model by 16.81% on the accuracy, 16.31% on precision, and 15.93% on recall. This demonstrates the strong ability of the model to achieve high performance on limited datasets. Furthermore, it is seen that as the number of training samples varies, the second-best performing model loses consistency, i.e., for $n = 200$, the Naive Bayes is the second-best-performing model, its performance decreases for $n = 300, 400, 500$, where the Ridge classifier and Random forest are seen performing better. This illustrates that no single machine learning model exhibits robust and consistent performance while training on limited data. On the contrary, the proposed model, which outperforms other models, demonstrates consistency on all subset configurations of limited data.

Exp-2. Model Performance Comparison on Severely Imbalanced Data Configuration. Most multi-omics datasets available suffer from high-class imbalance (example, $75\% - 25\%$ class distribution in METABRIC trial) since positive samples are difficult to obtain compared to the negative ones. Most machine learning and deep learning models exhibit a reduction in performance in high class-imbalance. To demonstrate the model's robustness to class imbalance, we compare the model performance on two dataset subset configurations: *first* with no class imbalance ($50\% - 50\%$ class distribution) and *second* with a severely skewed class distribution of $75\% - 25\%$. The obtained results (see Table 3) are compared with machine learning models trained on the same subset configuration. Here, precision is considered a better evaluation metric than accuracy alone [18]. It is seen that the proposed model has consistent performance (Δ Precision $= -1.86$) on both the class distributions, whereas all other models show a significant decrease in performance (large variation in Δ Precision observed).

Table 1. Comparative Results with state-of-the-art machine learning models on limited data subset configurations for 5−fold cross-validation. Here **QDA:** Quadratic Discriminant Analysis, **Ada:** Ada Boost Classifier, **K-NN:** K-Neighbours Classifiers, **LightBGM:** Light Gradient Boosting Machine, **LR:** Logistic Regression, **DT:** Decision Tree, **GBC:** Gradient Boosting Classifier, **LDA:** Linear Discriminant Analysis, **RF:** Random Forest, **SVM:** Support Vector Machine (Linear Kernel), **ET:** Extra-Trees Classifier, **NB:** Naive Bayes

Model	n =**200**				n =**300**				n =**400**				n =**500**			
	Acc	**Prec**	**Rec**	**F1**	**Acc**	**Prec**	**Rec**	**F1**	**Acc**	**Prec**	**Rec**	**F1**	**Acc**	**Prec**	**Rec**	**F1**
QDA	51.06	52.78	54.19	53.36	57.93	58.79	64.76	61.29	49.47	49.58	55.71	51.59	56.46	59.52	57.54	58.42
Ada	53.33	53.77	58.48	55.79	66.02	66.52	69.39	67.86	68.81	69.15	67.33	68.07	77.38	78.22	80.16	79.04
K-NN	54.71	61.73	35.14	43.23	61.23	70.45	44.46	53.63	60.95	69.33	40.66	50.72	59.03	71.97	41.37	51.81
LightGBM	61.88	63.60	62.38	62.39	64.13	65.98	64.72	65.04	71.69	70.59	74.68	72.27	77.08	79.70	77.44	78.31
LR	61.96	63.61	64.10	63.45	69.81	71.75	69.44	70.49	75.26	77.30	70.95	73.85	76.24	77.71	79.59	78.34
DT	61.98	65.97	63.90	64.15	63.61	64.47	65.63	64.49	70.60	71.88	68.10	69.57	72.19	74.09	74.74	74.23
Ridge	62.01	65.17	61.33	62.64	67.48	68.93	67.62	68.18	76.70	81.15	70.32	74.86	73.93	76.43	74.21	75.20
GBC	64.87	68.09	68.00	66.53	70.33	73.22	68.53	70.51	74.55	76.75	69.66	72.90	78.82	**81.11**	78.52	79.76
LDA	65.48	66.63	67.90	66.98	62.23	66.01	59.39	62.12	62.73	63.98	55.74	59.34	63.03	66.99	59.69	63.03
RF	66.24	72.59	62.57	65.75	67.44	69.60	66.62	67.57	70.95	71.79	68.10	69.72	78.82	80.47	80.11	80.20
SVM	66.27	70.69	66.57	67.09	66.49	66.83	71.30	68.76	73.10	71.00	78.23	73.97	74.24	76.87	76.34	76.23
ET	66.35	72.51	64.10	67.08	69.85	70.23	72.25	71.11	74.87	73.30	76.72	74.81	77.67	80.39	77.97	78.97
NB	73.41	73.68	**77.90**	75.01	69.86	69.40	74.98	71.70	68.81	68.15	70.26	68.65	74.23	74.04	79.54	76.53
Proposed	**77.50**	**85.0**	73.91	**79.07**	**86.67**	**85.71**	**90.91**	**88.24**	**83.75**	**83.33**	**81.08**	**82.19**	**83.00**	79.55	**81.40**	**80.46**

Table 2. Qualitative and Quantitative comparison of results on METABRIC trial data with proposed models in literature.

Model	Acc. ↑	Prec. ↑	Rec. ↑	F1 ↑	Dataset↓	Model Training ability data		
						Limited	Incremental	Imbalance
Deep Learning models								
Khademi et al., 2013 [19]	80.00	–	–	–	1980	Low	–	Low
Khademi et al., 2015 [4]	82.00	–	–	–	1980	Low	–	Low
Sun et al., 2019 [6]	82.60	74.90	45.00	–	1980	Low	–	Low
Arya et al., 2020 [7]	**90.20**	84.10	74.70	–	1980	–	–	Low
Machine Learning models								
Quadratic Disc. Analysis	53.22	53.36	53.88	53.46	1000	Low	–	Low
K-Neighbor Classifier	61.95	73.97	36.68	48.79	1000	Low	–	Low
Linear Discriminant Anal.	69.52	72.25	63.02	67.26	1000	Low	–	Low
Decision Tree	73.11	73.49	72.78	73.04	1000	Low	–	Low
Naive Bayes	74.10	83.48	60.16	69.88	1000	Low	–	Low
Random Forest	75.96	74.87	78.51	76.50	1000	Low	–	Low
Ridge Classifier	76.82	80.33	71.34	75.47	1000	Low	–	Low
Extra Trees Classifier	77.39	77.08	77.93	77.48	1000	Low	–	Low
Ada Boost Classifier	79.39	80.42	77.93	78.94	1000	Low	–	Low
Support Vector Machine	79.54	80.61	78.22	79.16	1000	Low	–	Low
Light Gradient Boosting	79.83	81.22	77.65	79.30	1000	Low	–	Low
Logistic Regression	80.83	82.43	78.51	80.31	1000	Low	–	Low
Gradient Boosting	81.26	81.29	81.37	81.30	1000	Low	–	Low
Proposed method	87.00	**93.01**	**89.26**	**91.09**	**1000**	**High**	√	**High**

Table 3. Comparative Results with state-of-the-art machine learning models for 5−fold cross validation on imbalanced data subset configurations ($n = 1000$).

Model	50%-50%				25%-75%				Δ Prec.
	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	
QDA	53.22	53.36	53.88	53.46	53.36	26.50	47.63	33.94	−26.86
K-NN	61.95	73.97	36.68	48.79	72.96	39.55	09.68	14.56	−34.42
LDA	69.52	72.25	63.02	67.26	65.81	36.49	48.86	41.69	−35.76
DT	73.11	73.49	72.78	73.04	71.40	42.00	35.32	38.02	−31.49
NB	74.10	83.48	60.16	69.88	53.36	30.71	68.21	42.28	−52.77
RF	75.96	74.87	78.51	76.50	79.26	71.98	30.71	42.13	−2.89
Ridge	76.82	80.33	71.34	75.47	73.53	47.91	47.17	47.37	−32.42
ET	77.39	77.08	77.93	77.48	78.55	65.69	32.37	42.80	−11.39
Ada Boost	79.39	80.42	77.93	78.94	78.12	59.49	48.30	52.63	−20.93
SVM	79.54	80.61	78.22	79.16	77.83	59.58	50.62	53.58	−21.03
LightGBM	79.83	81.22	77.65	79.30	77.26	57.25	40.38	47.18	−23.97
LR	80.83	82.43	78.51	80.31	80.11	64.03	48.86	55.23	−18.40
GBC	81.26	81.29	81.37	81.30	78.98	63.24	38.63	47.78	−18.05
Proposed	**87.00**	**93.01**	**89.26**	**91.09**	**86.00**	**91.15**	**89.93**	**90.54**	**−1.86**

Exp-3. Overall Model Performance Comparison. For an in-depth quantitative and qualitative comparison, $n = 1000$ was taken as the subset configuration. The proposed model was trained for short and long-term patient survival prediction following the conventional $80 - 20\%$ train-test split, with hyperparameters $\theta = 0.7263$ and $\gamma = 2$, found experimentally during model hyperparameter tuning and parametric study (see Exp. 04). The number of $\mathcal{H}$ formed during training was found to be 12. The proposed model is compared with state-of-the-art machine learning and deep learning models on the METABRIC trial data, as shown in Table 2. The obtained results show that the proposed model outperforms all other machine learning models by a significant margin ($\approx 5.74\%$ improvement on accuracy, $\approx 9.53\%$ on precision). Moreover, the model performance is at par with other deep learning approaches despite training on a limited dataset. It can be inferred that the proposed model can predict patient survival with comparable accuracy and higher precision while requiring a relatively small dataset to train. The primary reason for this is the data-intensive nature of the majority of proposed models, which fail when applied to imbalanced datasets. Moreover, the proposed model preserves both the contrasting and similar characteristics of each class sample. This differs from deep learning models that learn mostly the contrasting features while minimizing a loss function $\mathcal{L}$. The results confirm that the 'fuzzy' way is more suited to imbalanced and limited data.

Exp-4. Parametric Study and Time Complexity Analysis. A parametric study was performed to analyze the effect of variations in hyperparameters θ and γ. From Figure 3, we conclude that as θ increases, the number of hyperboxes ($\mathcal{H}$) created during training shows an 'exponential' increase rather than a 'linear' one. In contrast, the training time first shows a sharp rise until $\theta = 0.2$, after

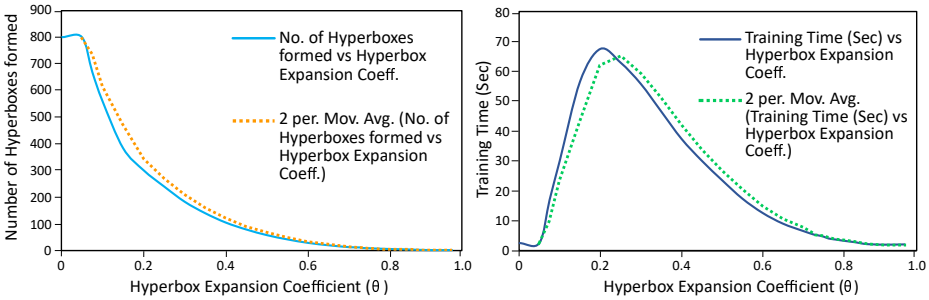


Fig. 3. Effect of variation in hyperbox ($\mathcal{H}$) expansion coefficient on **A.** Number of hyperbox ($\mathcal{H}$) formed during model training **B.** Model training time (*sec*)

which it decreases exponentially. The study quantifies that the proposed model has a significantly lower total training time (≈ 5 sec) compared to other models. However, the test time is comparatively higher (≈ 90 sec/ sample).

4 Discussion and Conclusions

The present study proposes a *de novo* approach towards the development of prognosis models, framing the task as an 'incremental learning' problem. The proposed model addresses the problem of the limited availability of high-throughput multi-omics datasets and the high-class imbalance seen in them. The obtained results establish the model's effectiveness and quantify that fuzzy classifier-based models are more suited towards problems where the dataset is highly imbalanced or limited, such as developing prognosis models combining multi-omics data. The proposed model surpasses state-of-the-art machine learning models significantly. These results suggest that prediction through 'fuzzy intelligence' is a promising approach towards breast cancer prognosis. In future work, we aim to expand the model for ovarian cancer, cervical cancer, fallopian tube cancers, etc., among others, caused by BRCA1 and BRCA2 mutations. Future research may also integrate a fourth multi-modal data, such as gene methylation, miRNA expression, or pathology image dataset, to improve classification performance further.

References

1. van de Vijver, M.J., He, Y.D., van't Veer, L.J., Dai, H., Hart, A.A.M., Voskuil, D.W., Schreiber, G.J., Peterse, J.L., Roberts, C., Marton, M.J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E.T., Friend, S.H., Bernards, R.: A gene-expression signature as a predictor of survival in breast cancer. N. Engl. J. Med. 347, 1999–2009 (2002).
2. Xu, X., Zhang, Y., Zou, L., Wang, M., Li, A.: A gene signature for breast cancer prognosis using support vector machine. In: 2012 5th International Conference on BioMedical Engineering and Informatics. IEEE (2012).

3. Nguyen, C., Wang, Y., Nguyen, H.N.: Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic. J. Biomed. Sci. Eng. 06, 551–560 (2013).
4. Khademi, M., Nedialkov, N.S.: Probabilistic graphical models and deep belief networks for prognosis of breast cancer. In: 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA). IEEE (2015).
5. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2015).
6. Sun, D., Wang, M., Li, A.: A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data. IEEE/ACM Trans. Comput. Biol. Bioinform. 16, 841–850 (2019).
7. Arya, N., Saha, S.: Multi-modal classification for human breast cancer prognosis prediction: Proposal of deep-learning based stacked ensemble model. IEEE/ACM Trans. Comput. Biol. Bioinform. 1–1 (2020).
8. Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., Altman, R.B.: Missing value estimation methods for DNA microarrays. Bioinformatics. 17, 520–525 (2001).
9. Gevaert, O., De Smet, F., Timmerman, D., Moreau, Y., De Moor, B.: Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. Bioinformatics. 22, e184-90 (2006).
10. Ding, C., Peng, H.: Minimum redundancy feature selection from microarray gene expression data. J. Bioinform. Comput. Biol. 03, 185–205 (2005).
11. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. Pattern Anal. Mach. Intell. 27, 1226–1238 (2005).
12. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Teh, Y.W. and Titterington, M. (eds.) Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. pp. 249–256. PMLR, Chia Laguna Resort, Sardinia, Italy (2010).
13. Simpson, P.K.: Fuzzy min-max neural networks. I. Classification. IEEE Trans. Neural Netw. 3, 776–786 (1992).
14. Alpern, B., Carter, L.: The hyperbox. In: Proceeding Visualization '91. IEEE Comput. Soc. Press (2002).
15. Gabrys, B., Bargiela, A.: General fuzzy min-max neural network for clustering and classification. IEEE Trans. Neural Netw. 11, 769–783 (2000).
16. Nandedkar, A.V., Biswas, P.K.: A General Reflex Fuzzy Min-Max Neural Network. Engineering Letters. 14, 195–205 (2007).
17. Curtis, C. et al.: The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 486, 346–352 (2012).
18. Saito, T., Rehmsmeier, M.: The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One. 10, e0118432 (2015).
19. Khademi, M.: Probabilistic graphical models for prognosis and diagnosis of breast cancer, (2014).