Prabha Materials Science Letters
Vol. 5, No. 1, 221-243, 2026
https://doi.org/10.33889/PMSL.2026.5.1.012

VAE-Guided Transformer Based Deep Learning Framework for Inverse Prediction of Aluminium (Al) Alloy Post-Processing Conditions

Harjit Pal Singh
Department of Physics,
The University of the West Indies, St. Augustine, Trinidad and Tobago.
E-mail: drhpsingh83@gmail.com

(Received on December 25, 2025; Revised on February 28, 2026; Accepted on Revised on February 28, 2026)

Abstract
These days, Machine Learning (ML) is widely used in material science research due to various advancements in algorithmic architecture. ML algorithms can be used for forward and inverse design prediction of the materials. Based on material compositions and processing, inverse design material informatics can produce any specific material profile. However, this machine learning approach is limited due to lack of data and presence of imbalanced datasets. The problem of imbalanced data severely impacts the outcome due to biased machine learning models that favor majority classes and ignore the rare classes in the datasets. To address these issues, the work in this paper proposed a comprehensive deep learning framework for inverse prediction for aluminium (Al) alloys using post-processing conditions. The proposed framework utilized the generative manifold-guided variational autoencoder (VAE) along with the deep learning-based TabTransformer for classification tasks. This transformer-based attention-aware VAE-guided framework worked well enough to capture complex interactions among compositional and mechanical features of Al alloy samples. The performance evaluation results demonstrated that significant improvement was achieved through the proposed framework of VAE-guided TabTransformer. The effectiveness of the proposed work was highlighted with a mass increment in macro F1-scores, which increased from 0.49 to 0.89 when compared to the baseline framework. The improvement was not only due to the majority classes, but the outcomes of minority classes also contributed significantly. Further, the performance was supported by K-fold and statistical validation results. The proposed framework successfully addressed the complex data-imbalanced inverse design problem and proved to be a powerful, generalizable, and intelligent tool for material discovery and optimization.

Keywords- Inverse prediction, VAE, Transformer, Machine learning, Deep learning.

1. Introduction

Metals and their alloys have an important role in today's modern technology and are widely used in many industries such as electronics, biomedical engineering, automotive, construction, and aerospace (Boyer, 1995; Magdassi et al., 2010; Zhang, 2011; Prasad & Wanhill, 2017; Wahid et al., 2020). Their capacity to adjust basic characteristics such as ductility, corrosion, conductivity, strength, and resistance, through alloying and further processing stages accounts for their exceptional adaptability. Particularly, aluminium (Al) alloys are significantly important for high-strength and lightweight applications and further, grouped into series 7xxx, 6xxx, and 2xxx based on their fundamental alloying elements and their processing procedures (Polmear, 2005; Bhat et al., 2024b). The nonlinear relationship between compositional features and thermochemical treatment increases the complexity of alloy designs such as precipitation hardening is dependent on the order and settings of the heat treatment steps that regulate the formation of nano-precipitates, which has a substantial impact on strength and ductility. The complex relationship between chemical processing, structure, and properties leads to an inverse design problem, which in return helps to find the best composition and processing method from the desired properties (Noh et al., 2019; Bang et al., 2024). The high multidimensional space and practical constraints require powerful data-driven algorithms that can explore and understand these heterogeneous landscapes.

The advancements in machine learning algorithms revolutionized material science research. With machine learning algorithms, the compositional and processing data based forward predictions for the new material design can be made through learned patterns and trends (Zivic et al., 2025). The prediction for classes and properties of the materials in metal and alloys can be made based on the chemical formula (Ramprasad et al., 2017; Butler et al., 2018; Jha et al., 2018; Zhuang & Barnard, 2023, 2024). The predictions through forward properties are limited in the selection of specific material for applications and post- processing conditions can only address the material design problems (Jain et al., 2013). It was observed that the complex inverse problem in material science, which requires high dimensional solution spaces, cannot be solved with traditional machine learning techniques (Liu et al., 2024). To predict the material and processing method matched with defined objectives for alloys, the inverse design methods utilize algorithms such as reinforcement learning, Bayesian optimization, surrogate modelling, and generative models (Lu et al., 2022; Wang et al., 2022; Han et al., 2025). However, due to imbalance, heterogeneity, and other noise in the data, it demands advanced data handling methods and algorithmic architectures to ensure interpretable, stable, and generalizable predictions. Further, the deep learning algorithms and architectures, such as graph neural networks, transformers and convolutional networks, are proven to be effective in capturing the complexities of high-dimensional data commonly found in materials datasets (Wang et al., 2022; Liu et al., 2024). The attention-enabled transformer-based algorithms can effectively process complex tabular datasets based on non-linear interactions to predict the relationships between features of the material alloys (Badaro et al., 2023; Attari & Arroyave, 2025). But the imbalance dataset is the major challenge in machine learning for materials since rare alloys or infrequent processing conditions are neglected, which in turn leads to biased classification toward majority classes (Jiang et al., 2025). The oversampling methods can balance the data by introducing synthetic data samples and solve the problem of imbalance dataset (Chawla et al., 2002; Gosain & Sardana, 2017; Sharma et al., 2022; Sharma & Gosain, 2023). The combination of the oversampled techniques with attention-based algorithms can efficiently manage the complexity and imbalance in tabular datasets of material alloys (Brishti et al., 2025; Madani et al., 2025; Wang et al., 2025a).

This work presents an inverse prediction framework for aluminium (Al) alloy based post-processing classifications. In this framework, the deep learning TabTransformer model was integrated with Variational Autoencoder (VAE) guided oversampling to address the problem of inverse prediction of Al alloys. The framework was processed with the publicly available dataset, which was heterogeneous and severely imbalanced. The advanced generative VAE oversampling of minority classes produced highly practical diverse data samples to balance the minority classes in the dataset and make it feasible to use TabTransformer classification tasks. The augmented dataset was further processed through the multi-head self-attention architecture to capture the non-linear interactions for effective prediction. The results of the performance evaluation of the proposed framework demonstrated that significant improvement was achieved in the classification for both minority and majority classes. The macro and weighted F1-scores analysis highlighted the efficacy of the proposed framework. The results were validated through statistical and K-fold validations. It was observed that the proposed framework successfully addressed the inverse alloy design problem by offering a scalable and data-efficient tool in material science design innovations.

2. Materials and Methods

This section was focused on methodology, data processing and the proposed detailed architectures of TabTransformer and VAE guided oversampling.

2.1 Methodology

The methodology for the work presented in this paper was focused on the inverse design prediction of materials based on post-processed conditions through VAE-guided oversampled TabTransformer deep

learning framework. The preprocessing was performed on the original imbalanced data samples and classes. Further, the inverse design evaluation was performed for baseline framework, where training and validation were performed on original imbalanced dataset and the proposed framework, where data was balanced through VAE guided manifolds. The K-fold and statistical validation were also performed to ensure performance. The whole process was divided as data processing, TabTransformer architecture, Guided Oversampling -Variational Autoencoder (VAE), performance evaluation and conclusion. The workflow was presented in **Figure 1**.

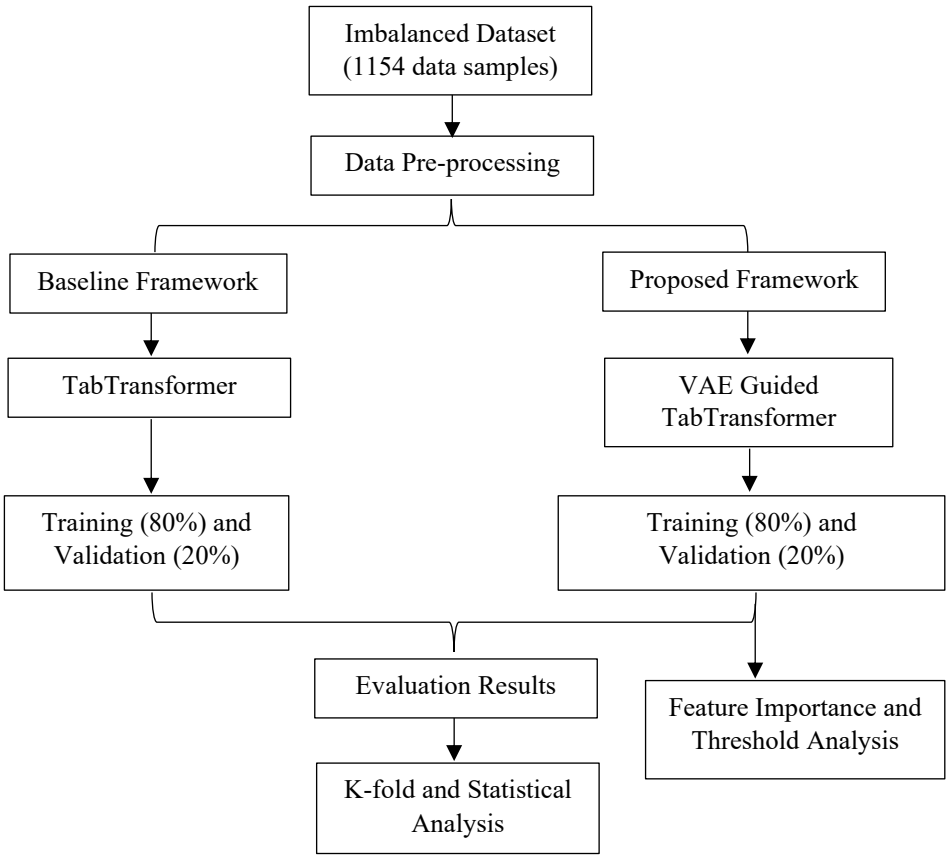


Figure 1. Workflow diagram VAE-guided TabTransformer framework.

2.2 Data Preprocessing

The analysis in this work was performed through an open-source aluminium (Al) alloy dataset (Bhat et al., 2023a, Bhat et al., 2023b; Barnard, 2024; Bhat et al., 2024a; Bhat et al., 2024b). This heterogeneous and imbalanced dataset contained 1154 individual alloy samples aggregated from various academic and industry sources (Bhat et al., 2024b). Each sample was described by a 28-dimensional continuous feature vector with categorical label indicating its processing condition. The feature space was designed to address the inverse problem, comprising 25 compositional and 3 mechanical features. The inclusion of all compositional and mechanical features reflected the better relationship between property, composition, and processing of materials, providing better metallurgical insight, and resulted in a feature space that was more extensive and identifiable for classification. The data was pre-processed to remove the outliers, artifacts, statistically insignificant classes/samples and samples corresponding to imprecise labels and to make it

suitable for further analysis. The classification task in this work was performed on a target variable-*Processing Condition* and the dataset was consolidated into eight significant and distinct metallurgical classes as No Processing (class 1), Solutionised + Artificially peak aged (class 2), Solutionised + Artificially over aged (class 3), Strain Hardened-Hard (class 4), Solutionised + Naturally aged (class 5), Strain hardened (class 6), Artificial aged (class 7) and Solutionised + Cold Worked + Naturally aged (class 8). The classes in the pre-processed dataset were severely imbalanced which could lead to inaccurate and biased results. The complexity due to imbalance, high dimensional feature space and non-linear interactions among compositional, processing and mechanical properties led to the development of deep learning powered VAE-guided oversampled TabTransformer framework to address the inverse prediction for material science.

2.3 TabTransformer Architecture for Inverse Prediction

The continuous input features were converted into categorical representations through discretization process (Badaro et al., 2023; Aurpa et al., 2024; Liu et al., 2024; Attari & Arroyave, 2025; Wang & Nguyen, 2025, Wang et al., 2025b). The discretization function $\{Y_j\}_{j=1}^{28}$ mapped a continuous feature $x_j \in R$ into one of $k = 10$ discreate bins using quantile-based binning and transformed the input vector into a categorical index vector as:

$$Y_j: R \to \{0, 1, \dots, k-1\} \quad (1)$$

$$c = [c_1, c_2, \dots, c_{28}], \text{ where, } c_j = Y_j(x_j) \quad (2)$$

At the embedding layer, columnar embedding was performed by maintaining a dedicated embedding matrix, $E_j \in R^{k \times d}$, for each discretized feature, where d was dimension of feature embedding. The corresponding embedding vectors were retrieved using categorical index, c_j, as:

$$E_j[c_j] \in R^d \quad (3)$$

The initial Transformer input matrix was formed by stacking embeddings of all 28 features as:

$$E^0 = \begin{bmatrix} E_1[c_1]^T \\ E_2[c_2]^T \\ \vdots \\ E_{28}[c_{28}]^T \end{bmatrix} \in R^{28 \times d} \quad (4)$$

The transformer encoder layer in the architecture contained M identical encoder layers and at each layer, $m \in \{1, \dots, M\}$, the input embeddings were transformed into output, $E^{(m)}$, through multi-head self-attention (MHA) and add & norm and Feed-Forward network (FFN).

For each attention head, $h \in \{1, \dots, H\}$, projection matrices, $P_{Q,h}, P_{K,h}, P_{V,h} \in R^{d \times d_k}$, linearly project the input to queries Q_h, keys K_h and values V_h respectively, where $d_k = \frac{d}{H}$:

$$Q_h = E^{(m-1)} P_{Q,h} \quad (5)$$

$$K_h = E^{(m-1)} P_{K,h} \quad (6)$$

$$V_h = E^{(m-1)} P_{V,h} \quad (7)$$

The attention weights were computed as:

$$head_h = softmax\left(\frac{Q_h K_h^T}{\sqrt{d_k}}\right) V_h \quad (8)$$

Concatenated output from all heads was as:

$$MHA\left(E^{(m-1)}\right) = Concat\ (head_1, head_2, \dots, head_H)P_o\ , P_o\ \in R^{d\times d} \quad (9)$$

During add & norm process, the MHA output was added to the input through the residual connection, followed by Layer Normalization:

$$A_{res} = LayerNorm(E^{(m-1)} + MHA\left(E^{(m-1)}\right)\ ) \quad (10)$$

The position-wise FFN was processed at the resulting A_{res} as:

$$FFN(A_{res}) = ReLU(AP_1\ + b_1)\ P_2\ + b_2 \quad (11)$$

The final output of this layer was obtained through another Add & Norm step:

$$E^{(m)} = LayerNorm(A_{res} + FFN(A_{res})) \quad (12)$$

Further, the feature embeddings, $E^{(m)}$, were flattened into a vector, $v_{flat} \in R^{28d}$, followed by multi-layer perception (MLP) to get logits, $z \in R^8$:

$$z = MLP\ \left(flatten\left(E^{(m)}\right)\right) \quad (13)$$

After the application of SoftMax, the predicted class probabilities were computed as:

$$\hat{y} = softmax(z) \quad (14)$$

The architecture of TabTransformer model was presented in **Figure 2** and the hyperparameters were presented in **Table 1**.

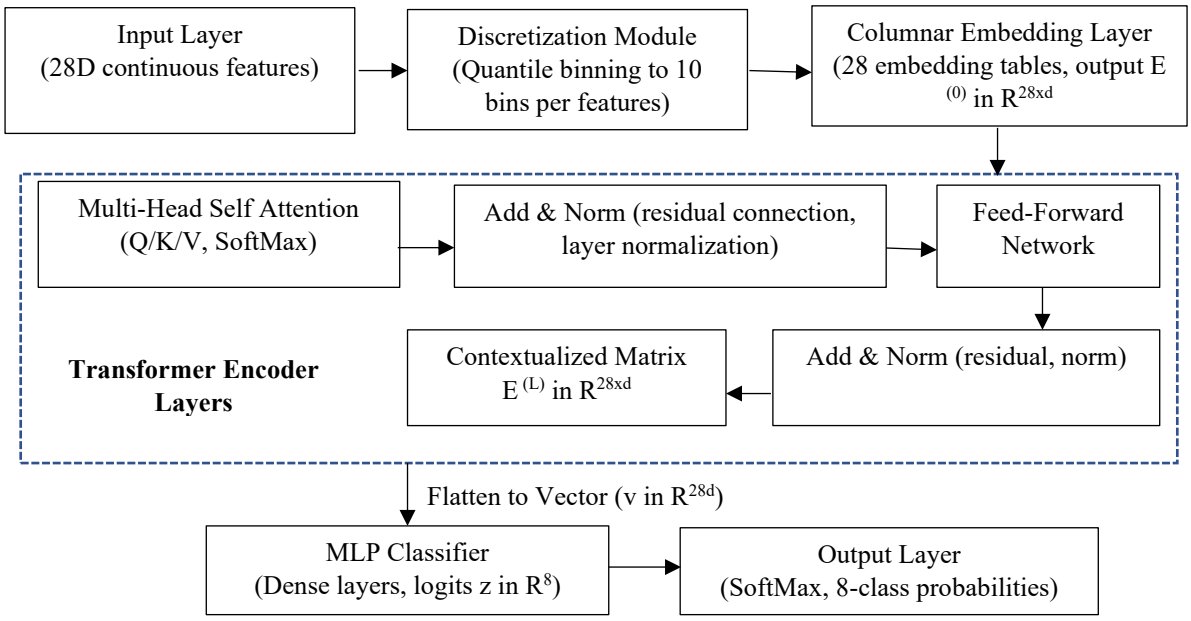


Figure 2. TabTransformer architecture.

2.4 Guided Oversampling -Variational Autoencoder (VAE)

Due to the problem of imbalanced dataset, the work in this paper presented a Variational Autoencoder (VAE) based guided generative oversampling framework. This oversampling framework learned a significant, nonlinear latent representation of the alloy dataset and effectively reflected the inherent data manifold features of practical alloy compositions. The synthetic minority class samples generated from the learned latent space were not only diverse but also consistent with physicochemical characteristics and fundamental structures of the materials (Ai et al., 2023; Lee, 2023; Tang et al., 2025). The VAE decoder generated precise synthetic data points through the extraction of samples from the latent manifold, and these data points were more realistic variations in minority classes than classic synthetic oversampling algorithms SMOTE and ADASYN (Chawla et al., 2002; Gosain & Sardana, 2017; Sharma et al., 2022; Sharma & Gosain, 2023; Barnard, 2024).

Table 1. Hyperparameters for TabTransformer.

Embedding Dimensions (d)	64
Transformer Layers (L)	4
Attention Heads (H)	8
FFN Hidden Dimensions	128
MLP Head Layers	2 (64, 32)
MLP Activation	ReLU
Optimizer	AdamW
Learning Rate	1.00E-04
Batch Size	64
Loss Function	Cross-Entropy Loss
Discretization Method	Quantile Binning
Discretization Bins (K)	10

This adopted method not only improved the model generalization but also enhanced the prediction of the alloy classification. By learning a probabilistic mapping from a high-dimensional data space $X \in R^{28}$ to a lower-dimensional continuous latent space $Z \in d_z$, the trained VAE acted as an expert and a deep generative model in guided oversampling process. The architecture of VAE included encoder/decoder formation structured with fully connected layers. With training on majority classes, VAE learned robust latent representation of vital data manifold which enabled the model to capture the practical correlations between mechanical and elemental alloy features. This manifold further guided the subsequent generations. For the minority class augmentation, the following procedure was performed until all the minorities were oversampled to the size of the majority classes and a balanced dataset was obtained.

Step-1 Selection of random two distinct data samples from minority class, C_{min}.

Step-2 Obtaining latent distribution parameters for these two samples through trained encoder as $(\mu_{z_a}, \log \sigma_{z_a}^2) and\ (\mu_{z_b}, \log\ \sigma_{z_b}^2)$.

Step-3 Generation of latent vectors using the reparameterization as

$$\mathrm{z} = \mu_{\mathrm{z}} + \sigma_{\mathrm{z}} \odot \epsilon, \text{ where, } \epsilon \sim \mathcal{N}(0, \mathrm{I}) \quad (15)$$

Step-4 Perform interpolation on learned manifold through interpolation coefficient, $\lambda \sim U(0,1)$ and compute new latent vector z_{new} with

$$\mathrm{z}_{new} = \lambda \mathrm{z}_a + (1 - \lambda)\mathrm{z}_b \quad (16)$$

Due to structured and smoothness of latent space, this new latent vector was an intermediate point on learned manifold of plausible alloy, not an average in the feature space.

Step-5 Generation of new synthetic data sample in 28D feature space by passing through the trained decoder as:

$$x_{new} = \text{Decoder}(z_{new}) \quad (17)$$

Step-6 Assign the minority class label C_{min} to the newly generated sample x_{new}.

The VAE training maximized the evidence lower bound (ELBO) on log-likelihood of the data points led to minimizing mean squared error. The VAE pipeline was presented in **Figure 3**.

3. Performance Evaluation

The performance of the classification task was evaluated for two sets of frameworks as baseline TabTransformer, and proposed VAE oversampled TabTransformer. Both the frameworks were trained and tested in a Tensorflow/Keras environment on machine with Intel(R) Core (TM) i9-13900HX, 32 GB RAM, NVIDIA GeForce RTX 4080 GPU. The data was stratified-split into training (80%) and validation (20 %) set.

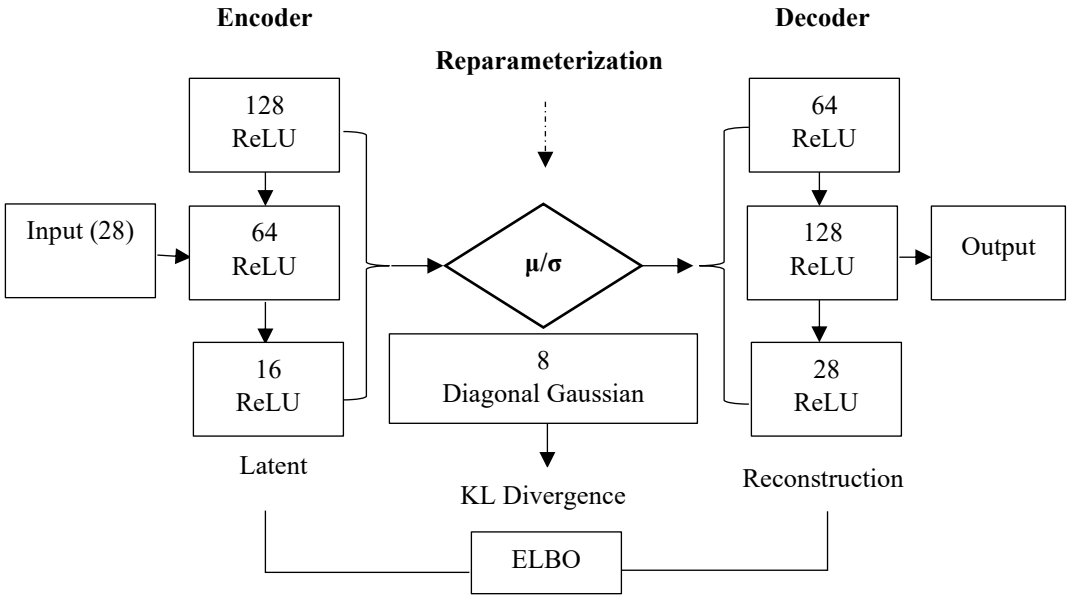
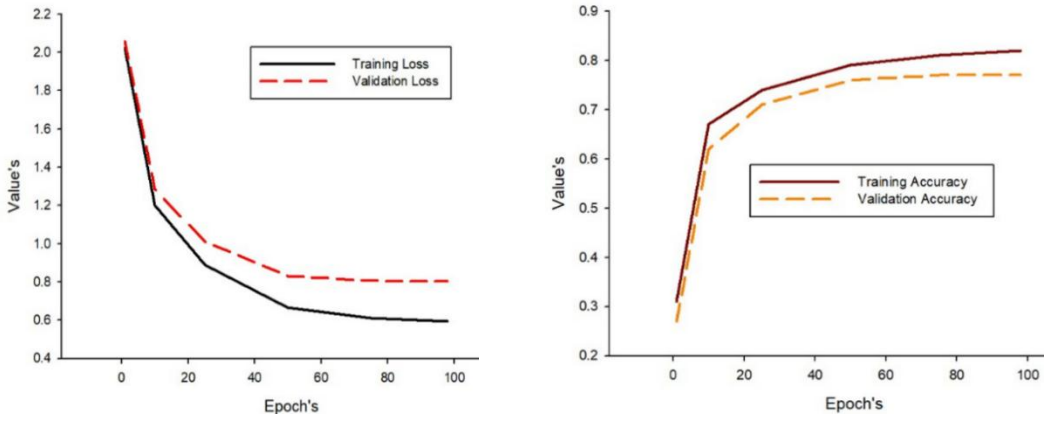


Figure 3. VAE architecture and training pipeline.

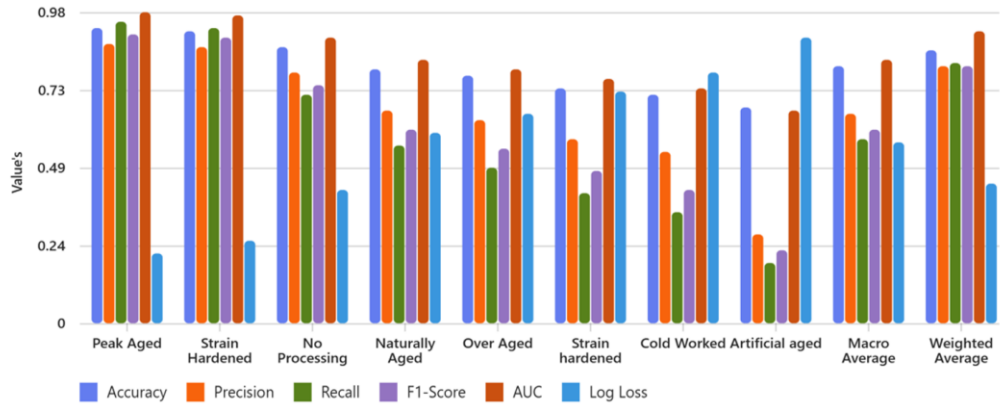
Case I-Baseline TabTransformer Framework

In this framework, the original imbalanced dataset was used to train the classifier, and further, validation was also performed on original dataset. The analysis of training and validation loss was presented in **Figure 4**. **Figure 5** presented the final performance metrics such as accuracy, AUC, precision, F1-score, recall and log-loss for baseline framework corresponding to each class. The results in **Figure 6** and **Figure 7** highlighted the F1-scores for each class for baseline framework training and validation. **Figure 8** presented clear distinction on performance of baseline framework through macro average and weighted average F1-score. The performance of the metrics precision, recall and AUC for baseline validation was presented in **Figure 9**.

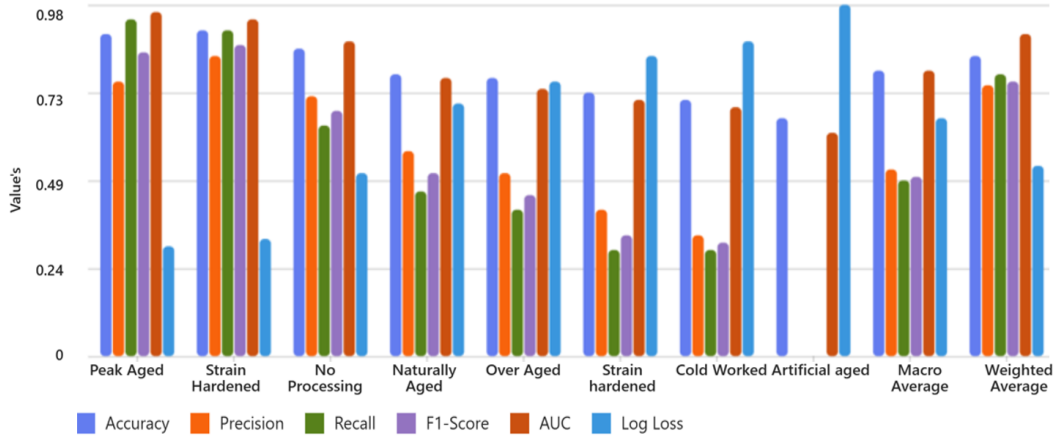


(a) (b)

Figure 4. Baseline (a) Training vs. validation loss (b) Training vs. validation accuracy.



(a)



(b)

Figure 5. Final performance metrics for baseline framework (a) Training (b) Validation.

(a) Discussion

At first glance, it was observed from the epoch-wise accuracy and loss analysis results from **Figure 4** that baseline framework worked consistently well as training and validation losses decreased with epochs and sufficient accuracies of 82% for training and 77% for validation were achieved. However, the breakdown of the F1-score trajectories and per-class performance metrics in **Figure 5** to **Figure 8** revealed the deeper insights of the baseline framework. The difference between weighted average and macro average F1 scores for both training and validation in baseline framework clearly highlighted the performance inconsistency among classes. For training, the baseline framework achieved 0.81 weighted average F1-score and the macro average value lagged to 0.61. However, this trend was even worse for validation results, where weight F1 score was 0.76 and macro F1 collapsed to 0.51 (0.518 ± 0.013, 10-fold) as presented in **Figure 8**. This led to the observation that generalization capacity of baseline model was dominated by major classes only at the cost of minority classes. The performance analysis of class-wise metrics revealed that the baseline framework delivered significant precision, accuracy, recall and F1 score for majority classes such as Peak Aged (class 2) and Strain Hardened (class 4) but for minority classes such as Cold Worked (class 8) recall observed value was very low as of 0.29 as presented in **Figure 5** and **Figure 6** and for Artificial aged (class 7), the baseline framework completely failed as shown in **Figure 9(b)** and F1-score as presented in **Figure 6** and **Figure 7**. Similar deficits trends were found across AUC, recall, precision and log-loss for minority classes as presented in **Figure 5** and **Figure 9**. These observations provided a clear, quantifiable benchmark which supported the need for an advanced and physically significant oversampling method to achieve fair and generalized prediction in classification tasks.

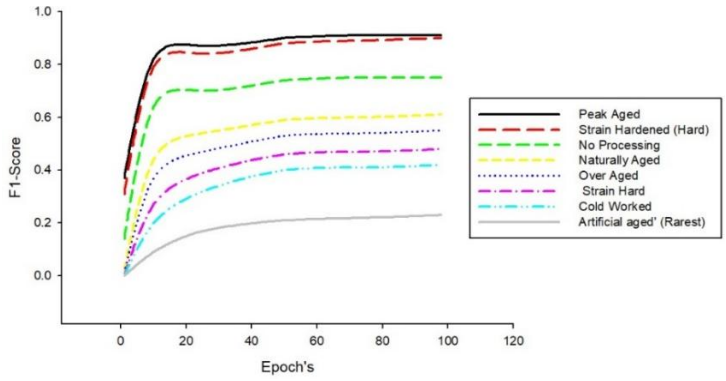


Figure 6. Classwise baseline training F1-score.

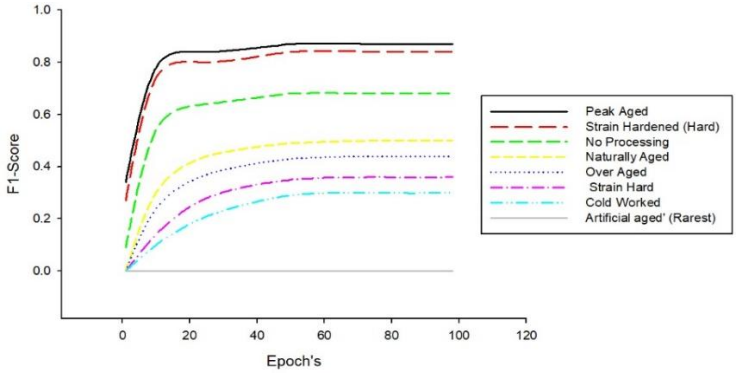


Figure 7. Classwise baseline validation F1-score.

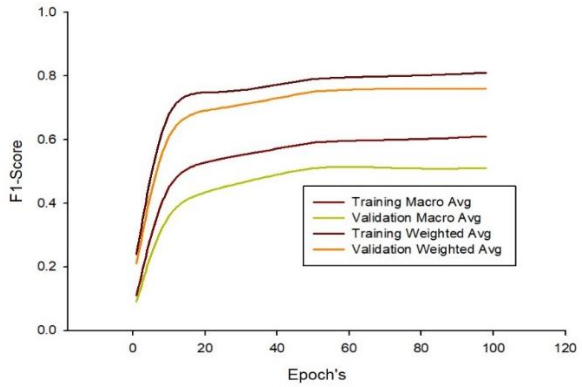


Figure 8. Baseline training/validation macro and weighted F1-score.

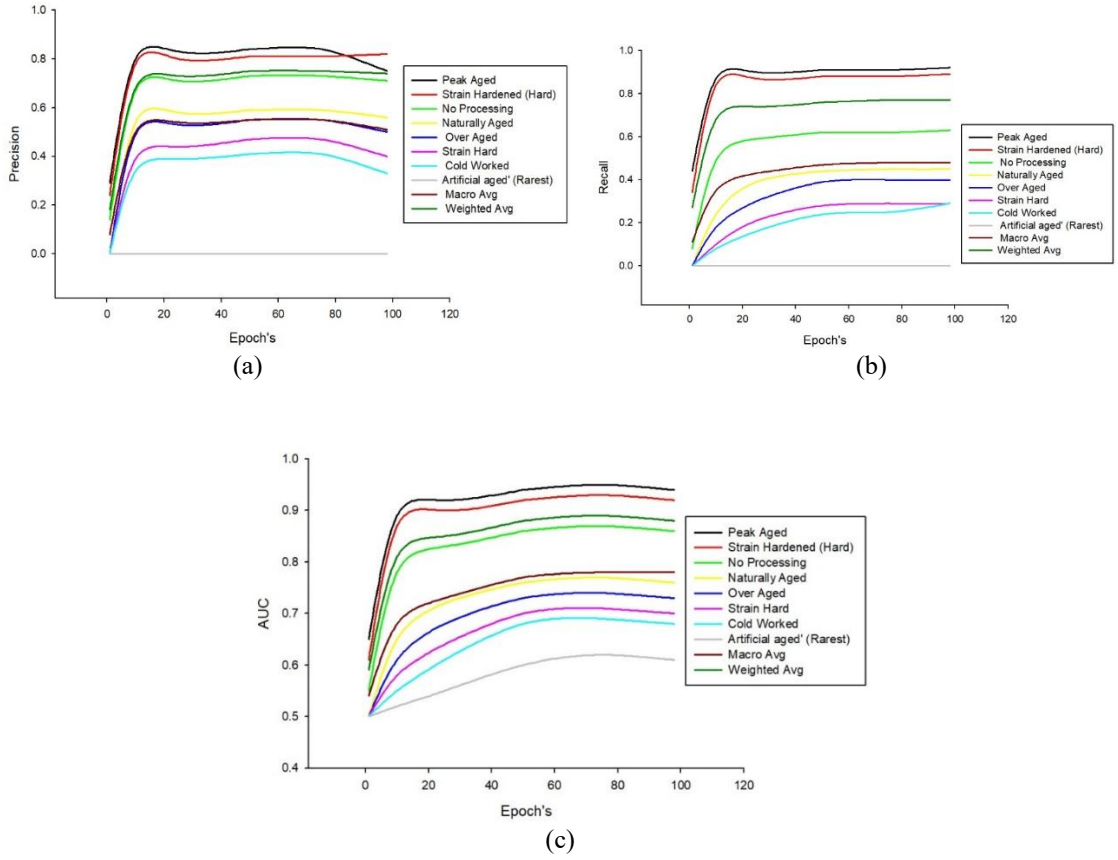
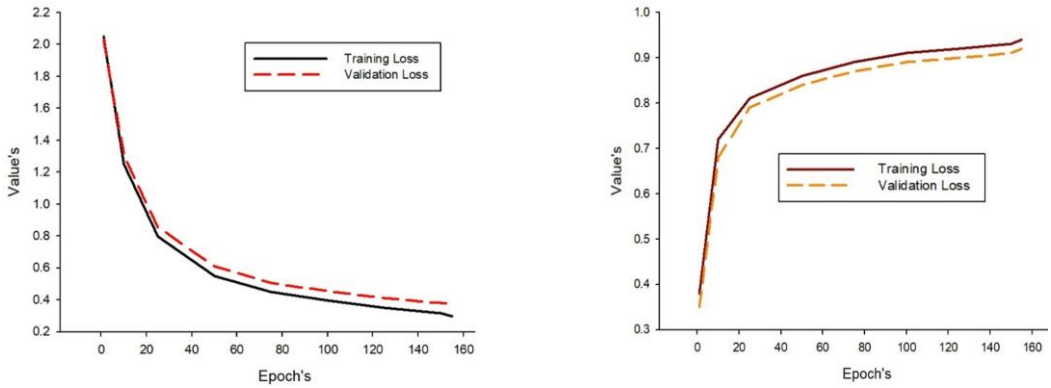


Figure 9. Baseline valdiation (a) Precision (b) Recall (c) AUC.

Case II- Proposed Framework

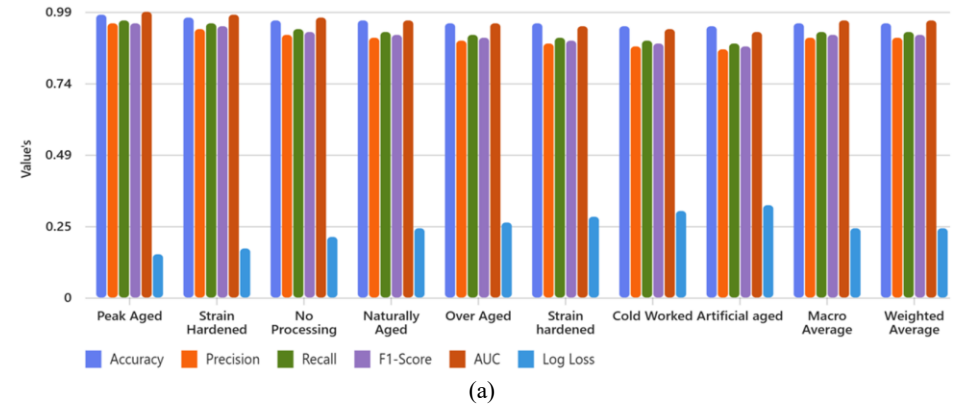
The proposed framework used for inverse prediction of Al alloy utilized VAE-guided for oversampling of the data samples to balance the original dataset. The performance of this proposed framework demonstrated statistically outstanding improvement in performance metrics and fairness in classification task, thus validating the effectiveness of the proposed deep learning TabTransformer framework. **Figure 10** presented

the loss and accuracy trade off with number of epochs for training and validation of the proposed framework. The class-wise final performance metrics for F1-score, precision, AUC, accuracy, recall and log-loss were presented in **Figure 11** for training and validation of the proposed framework. The training and validation F1-socres for each class were shown in **Figure 12**. The macro and weighted F1-socre analysis was presented in **Figure 13**. The validation precision, AUC and recall for the proposed framework were presented in **Figure 14**.

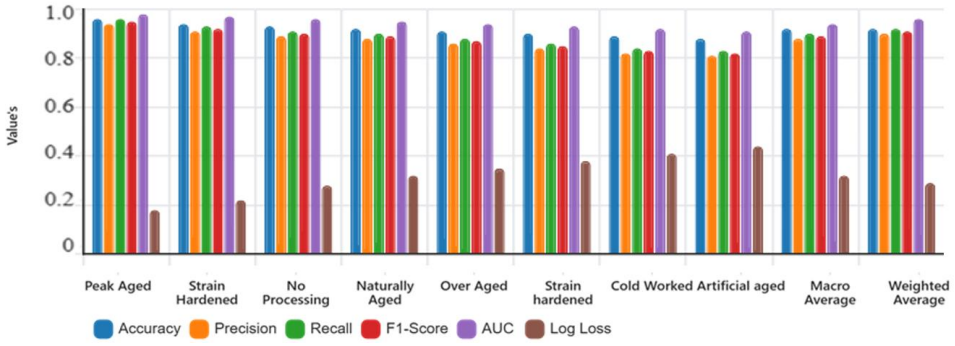


(a) (b)

Figure 10. Proposed framework (a) Training /validation loss (b) Training/validation accuracy.



(a)



(b)

Figure 11. Final performance metrics for proposed framework (a) Training (b) Validation.

(a) Discussion

The proposed VAE-guided oversampling deep learning framework demonstrated significant efficacy and robustness for multiclass Al alloy post processing conditioned dataset. The performance results of training and validation confirmed the effectiveness of the proposed framework to classify all eight classes efficiently with uniform proficiency. The early convergence and eventual stabilization of training and validation losses in **Figure 10(a)** indicated an effective optimization with minimum overfitting. For unseen and imbalanced validation data samples, the validation loss reached a minimum of 0.371 at 155 epoch. This outcome was further supported by training/validation accuracy trends in **Figure 10(b)**. The training accuracy was found to be 0.94 and the validation accuracy was approximately 0.92 with the proposed framework. Further, the final performance metrics at epoch 155 for each class in **Figure 11** and **Figure 14** revealed that the proposed multiclass framework achieved significantly high F1-score, accuracy, recall, AUC and precision for both majority and minority classes. It was evident from the outcome of **Figure 11** and **Figure 12** that the proposed framework was significantly robust and generalized to identify the most confronting minority classes. The major increment in F1-scores for Artificial Aged (class 8), Cold Worked (class 7) and Strain Hardened (class 6) were found to be 0.82, 0.83 and 0.85 respectively. The most crucial outcome of this framework was the almost overlapping of macro and weighted F1-scores for training and validation as presented in **Figure 13**, which firmly confirmed the excellent generalization capability of VAE guided oversampled TabTransformer based deep learning framework for inverse prediction of multiclass AI alloy. This framework excellently reconciled the class imbalance in the dataset and achieved outstanding prediction performance.

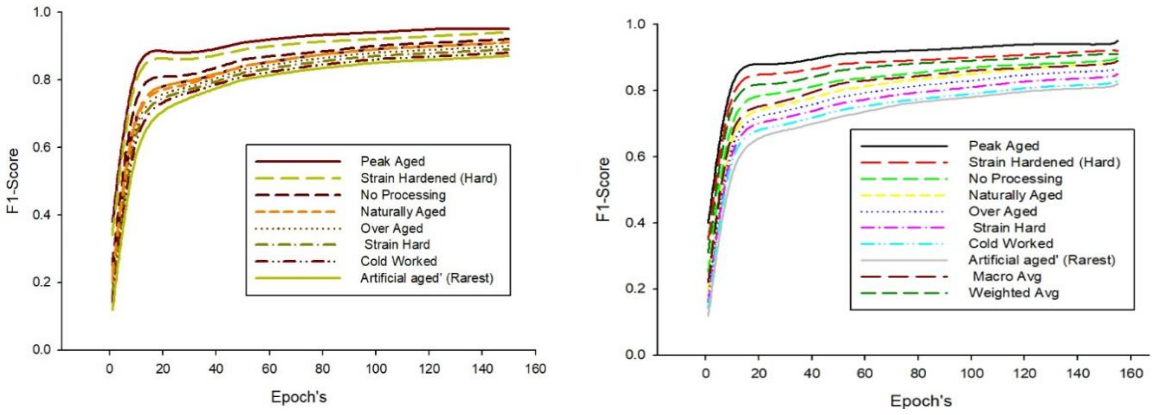


Figure 12. Class-wise F1-score for proposed framework (a) Training (b) Validation.

(b) K-fold Validation

The reliability and stability of the proposed framework was validated through K-fold (10-fold) validation technique. The results of 10-fold validation were presented in **Table 2**. The stable and high macro-average F1-score with mean of 0.887 was found to be the most significant outcome of this K-fold validation technique. Irrespective of which validation data was utilized, the consistency of the proposed framework was proven to be superior, and it efficiently learned to classify all chosen classes with excellent accuracy and fairness. The value of the standard deviation was found to be ±0.011, which was extremely low, and this demonstrated that the efficacy of this proposed work was not an artifact of favorable data split but rather a stable and recurring outcome of the adopted process. Additionally, it was observed that macro and weighted F1-scores, 0.887 and 0.907 respectively, were closely aligned. It was confirmed through these

results that proposed framework successfully learned the characteristics of both majority and minority classes. Further, the outcome of K-fold validation was also supported by high values of macro recall, macro-AUC and macro precision.

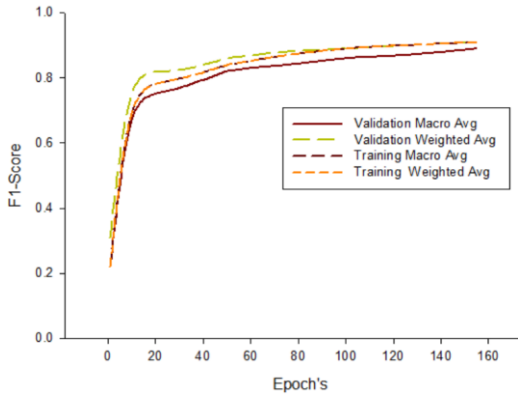


Figure 13. Proposed framework training/validation macro and weighted F1-score.

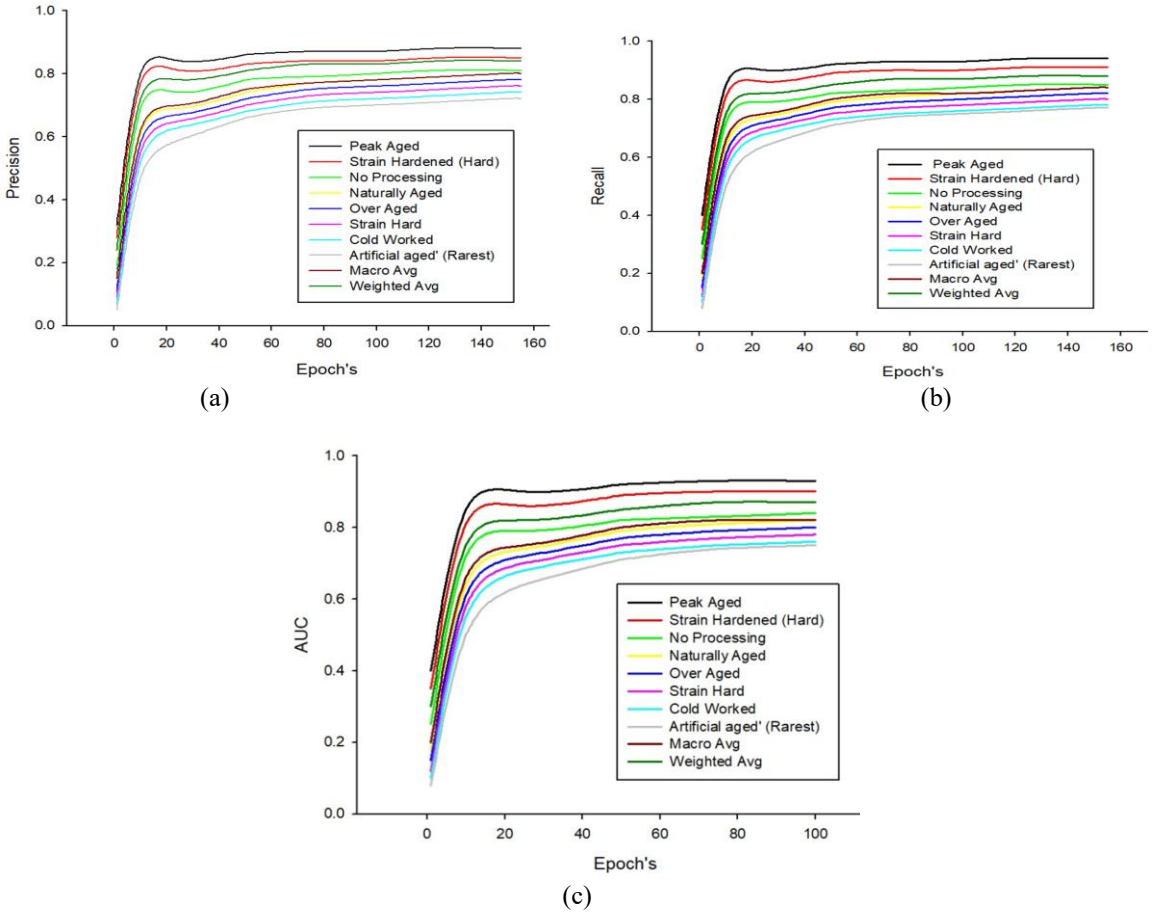


Figure 14. Proposed framework valdiation (a) Precision (b) Recall (c) AUC.

Case III- Feature Importance Analysis

In this section, the feature importance analysis was presented for baseline and proposed frameworks. The two-level analysis was performed using Shapley additive explanation (SHAP) based attribution method as global and class-wise, and the results were presented in **Figure 15** and **Figure 16**. The global feature importance profile averaged the contribution of a feature across all predictions as presented in **Figure 15**. The significant changes were observed in the model's learning priorities while analyzing the feature importance for both frameworks. The high concentration of the baseline framework was revealed towards the top features of the majority classes, but the minority classes were ignored. In contrast to baseline framework, in the proposed work the fair distribution of importance for majority and minority was highlighted. The dominance of the top features remained stable but the overall importance scores were significantly shifted towards the minority classes. The significant increment in the importance of the features such as Scandium (Sc) and Manganese (Mn) in the minority classes was observed in the proposed framework. The analysis of the importance of global features clearly validated the ability of the framework to understand the contextual features of mechanical properties and compositional elements. Further, the class-wise results provided a deep analysis of the feature importance for inverse prediction based on metallurgically relevance for each class in material science. Similar trends as of global analysis were observed for class-wise feature importance in baseline framework as presented in **Figure 16**. In comparison to baseline, the proposed framework again acted smart with a dynamic and context-aware intelligence. It was observed from **Figure 16(a)** that the importance scores for aluminium (Al) and elongation were predicted with significant accuracy. The dominant prediction was made for scandium (Sc) and Zirconium (Zr) in Artificial Aged (class 7) with importance scores of 22.5 % and 18.9 % respectively. These two predicted features were metallurgically important for this class of alloy in precipitation hardening and their ranking elevation supported the generalizability and reliability of the proposed framework.

Case IV- Ablation Study: Guided Threshold Variation

This ablation study was conducted to analyze the effect of guided threshold tuning in the proposed framework on classification performance for Al alloy processing category. The class-wise surface plots for threshold, recall and F1-score with threshold ranging from 2% to 6% were presented in **Figure 17**. It was observed from the results in **Figure 17** that the proposed framework mapped optimum relation between recall and F1-score at 6% threshold value, which, further, suggested that the proposed approach suppressed noise efficiently at 6% with significant accuracy and model sensitivity. The results of the majority classes such as Peak Aged and No Processing were high across a plateau of threshold points and led to the stability of the framework. For classes such as Cold Worked and Artificial Aged, the surface plots highlighted sharper optima, leading to careful tuning of threshold values to avoid underfitting. Furthermore, the generalization of the proposed framework was confirmed through macro-averaged surface plot since macro-averaged recall and F1-score achieved optimum value at 6% of threshold. This importance-driven threshold variational study supported the proposed framework with high generalization, high sensitivity for noisy and imbalanced tabular datasets.

Table 2. K-Fold validation performance.

Fold / Metric	1	2	3	4	5	6	7	8	9	10	Mean ± SD
Macro F1	0.88	0.90	0.87	0.89	0.88	0.90	0.89	0.87	0.90	0.89	0.887 ± 0.011
Weighted F1	0.90	0.92	0.89	0.91	0.90	0.92	0.91	0.89	0.92	0.91	0.907 ± 0.011
Accuracy	0.91	0.93	0.90	0.92	0.91	0.93	0.92	0.90	0.93	0.92	0.917 ± 0.011
Macro Precision	0.87	0.89	0.86	0.88	0.87	0.89	0.88	0.86	0.89	0.88	0.877 ± 0.011
Macro Recall	0.89	0.91	0.88	0.90	0.89	0.91	0.90	0.88	0.91	0.90	0.897 ± 0.011
Macro AUC	0.93	0.95	0.92	0.94	0.93	0.95	0.94	0.92	0.95	0.94	0.937 ± 0.011

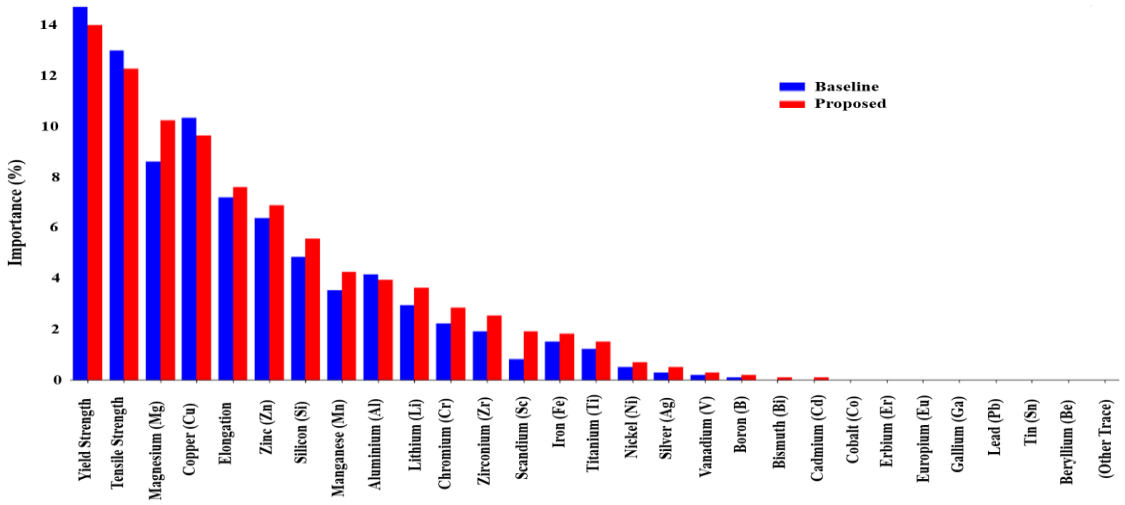
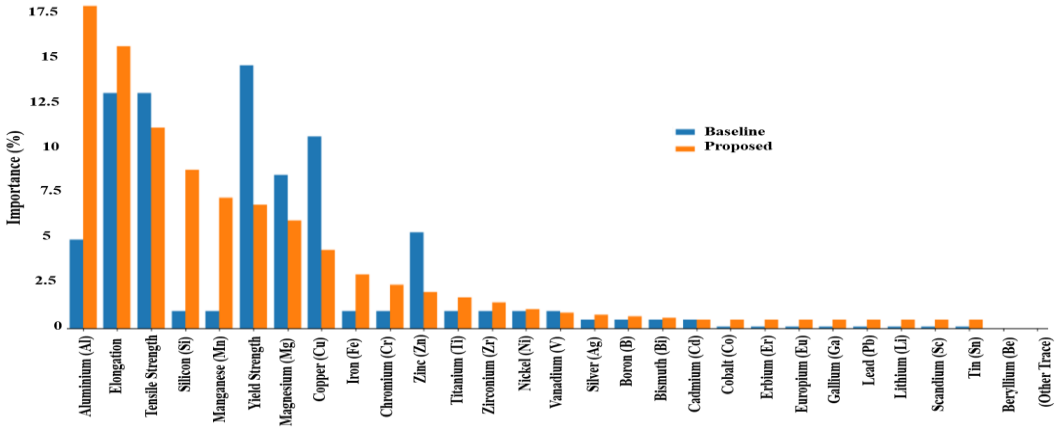


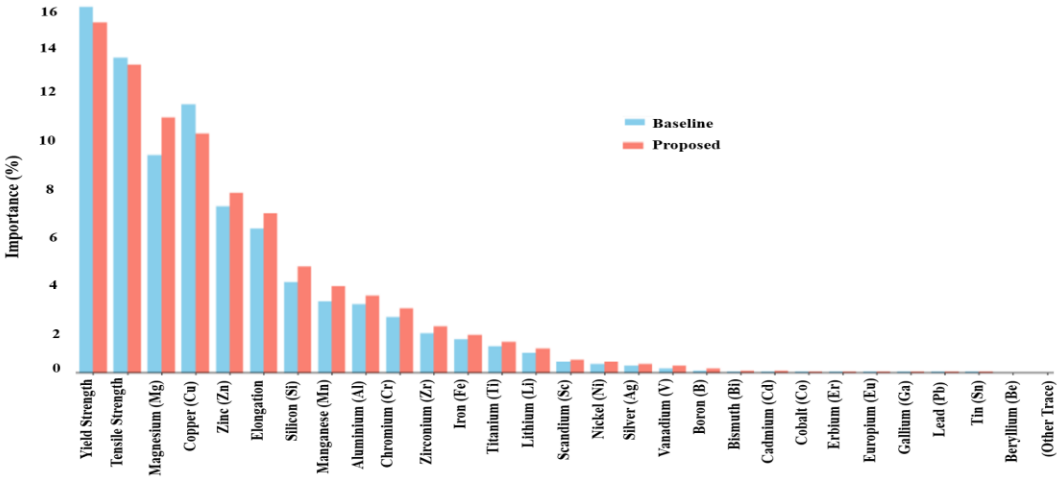
Figure 15. Global feature importance profile.

4. Statistical Validation

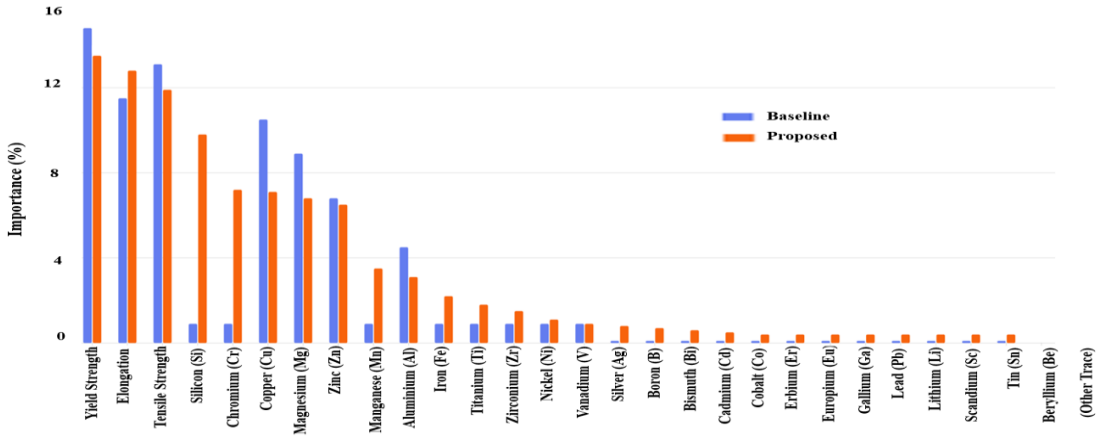
The statistical analysis was adopted to ensure the validation of the performance of the proposed VAE-oversampled framework. The one-tailed paired t-test was conducted for 10-fold resulted macro-average F1-scores of the baseline and the proposed frameworks. For this analysis, the null hypothesis was considered that mean macro F1-scores for both frameworks were same, and the alternative hypothesis assumed that mean macro F1-score of the proposed framework was significantly more than the baseline. The mean macro F1-scores for baseline and proposed frameworks were found to be 0.518 ± 0.013 and 0.887 ± 0.011 respectively. It was observed that the calculated p-value was less than 0.0001 ($p<0.0001$), below the $\alpha = 0.05$ threshold, which led to acceptance of alternative hypothesis by rejecting the null hypothesis. The high statistical confident results of the statistical analysis demonstrated that the increment in the performance of the proposed framework was significant and genuine rather than randomness. The low score variance confirmed the reliability and robustness of the proposed approach and supported the effectiveness of the proposed framework.



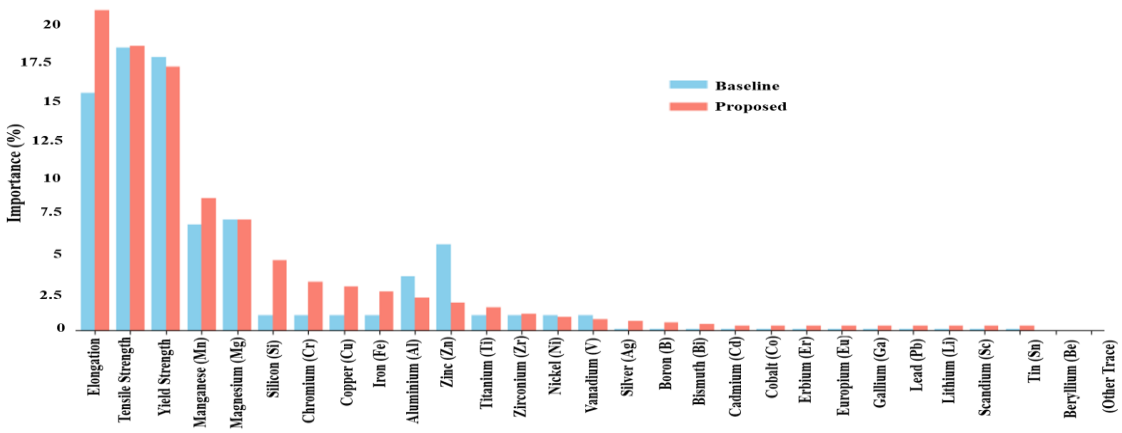
(a)


Importance (%)
Baseline
Proposed
Yield Strength
Tensile Strength
Magnesium (Mg)
Copper (Cu)
Zinc (Zn)
Elongation
Silicon (Si)
Manganese (Mn)
Aluminium (Al)
Chromium (Cr)
Zirconium (Zr)
Iron (Fe)
Titanium (Ti)
Lithium (Li)
Scandium (Sc)
Nickel (Ni)
Silver (Ag)
Vanadium (V)
Boron (B)
Bismuth (Bi)
Cadmium (Cd)
Cobalt (Co)
Erbium (Er)
Europium (Eu)
Gallium (Ga)
Lead (Pb)
Tin (Sn)
Beryllium (Be)
(Other Trace)

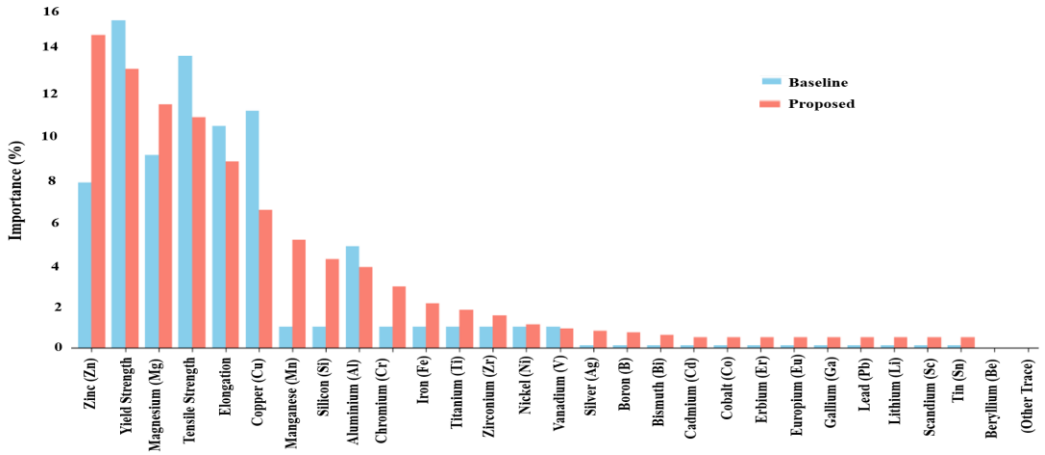
(b)


Importance (%)
Baseline
Proposed
Yield Strength
Elongation
Tensile Strength
Silicon (Si)
Chromium (Cr)
Copper (Cu)
Magnesium (Mg)
Zinc (Zn)
Manganese (Mn)
Aluminium (Al)
Iron (Fe)
Titanium (Ti)
Zirconium (Zr)
Nickel (Ni)
Vanadium (V)
Silver (Ag)
Boron (B)
Bismuth (Bi)
Cadmium (Cd)
Cobalt (Co)
Erbium (Er)
Europium (Eu)
Gallium (Ga)
Lead (Pb)
Lithium (Li)
Scandium (Sc)
Tin (Sn)
Beryllium (Be)
(Other Trace)

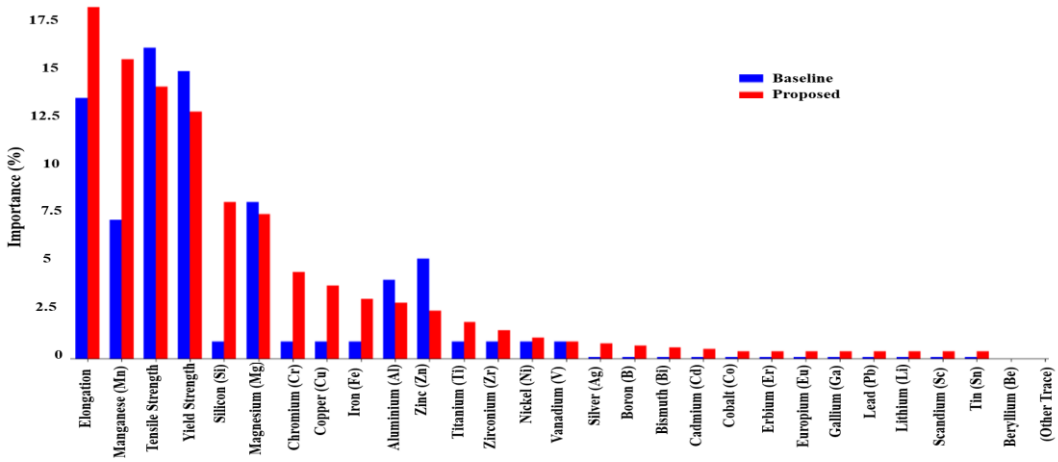
(c)


Importance (%)
Baseline
Proposed
Elongation
Tensile Strength
Yield Strength
Manganese (Mn)
Magnesium (Mg)
Silicon (Si)
Chromium (Cr)
Copper (Cu)
Iron (Fe)
Aluminium (Al)
Zinc (Zn)
Titanium (Ti)
Zirconium (Zr)
Nickel (Ni)
Vanadium (V)
Silver (Ag)
Boron (B)
Bismuth (Bi)
Cadmium (Cd)
Cobalt (Co)
Erbium (Er)
Europium (Eu)
Gallium (Ga)
Lead (Pb)
Lithium (Li)
Scandium (Sc)
Tin (Sn)
Beryllium (Be)
(Other Trace)

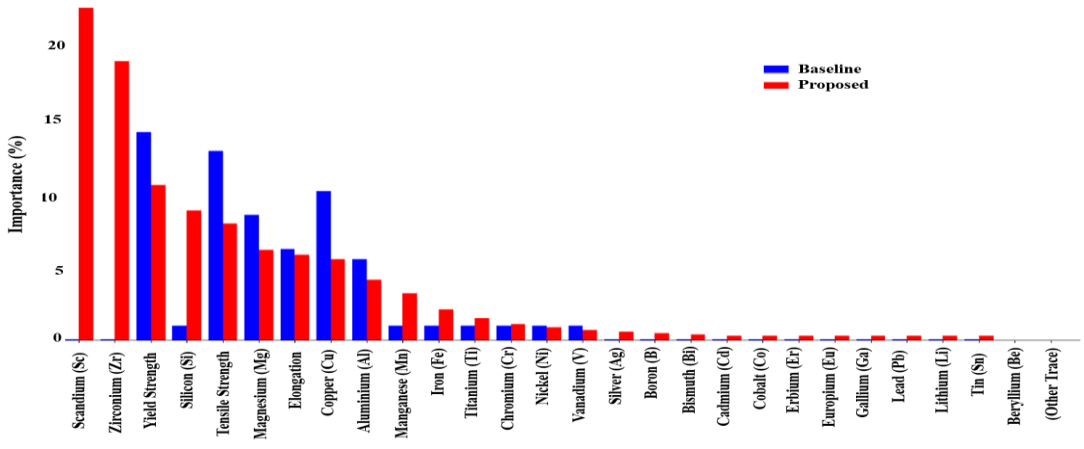
(d)


Importance (%)
Baseline
Proposed
Zinc (Zn)
Yield Strength
Magnesium (Mg)
Tensile Strength
Elongation
Copper (Cu)
Manganese (Mn)
Silicon (Si)
Aluminium (Al)
Chromium (Cr)
Iron (Fe)
Titanium (Ti)
Zirconium (Zr)
Nickel (Ni)
Vanadium (V)
Silver (Ag)
Boron (B)
Bismuth (Bi)
Cadmium (Cd)
Cobalt (Co)
Erbium (Er)
Europium (Eu)
Gallium (Ga)
Lead (Pb)
Lithium (Li)
Scandium (Sc)
Tin (Sn)
Beryllium (Be)
(Other Trace)

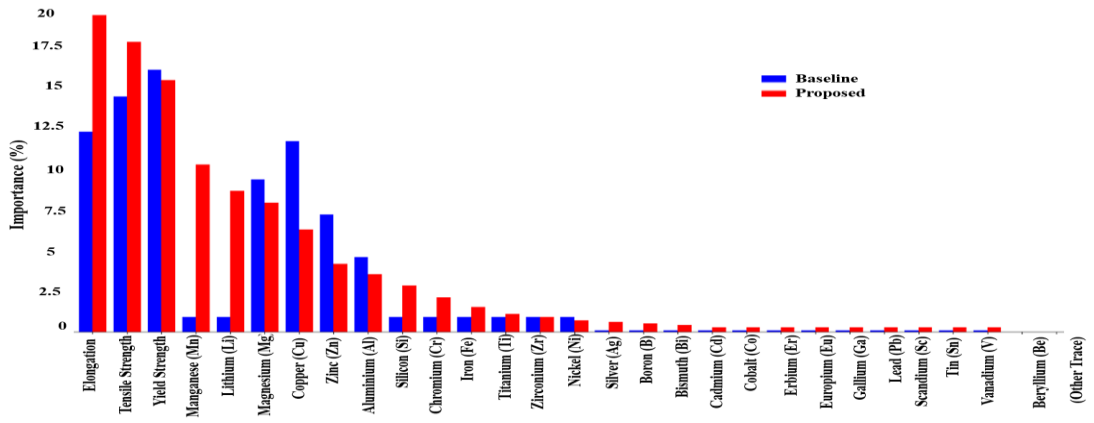
(e)


Importance (%)
Baseline
Proposed
Elongation
Manganese (Mn)
Tensile Strength
Yield Strength
Silicon (Si)
Magnesium (Mg)
Chromium (Cr)
Copper (Cu)
Iron (Fe)
Aluminium (Al)
Zinc (Zn)
Titanium (Ti)
Zirconium (Zr)
Nickel (Ni)
Vanadium (V)
Silver (Ag)
Boron (B)
Bismuth (Bi)
Cadmium (Cd)
Cobalt (Co)
Erbium (Er)
Europium (Eu)
Gallium (Ga)
Lead (Pb)
Lithium (Li)
Scandium (Sc)
Tin (Sn)
Beryllium (Be)
(Other Trace)

(f)

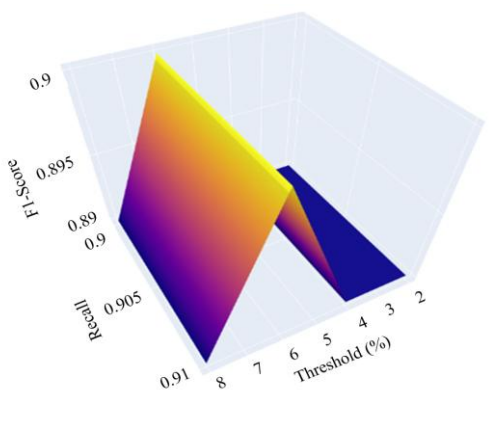

Importance (%)
Baseline
Proposed
Scandium (Sc)
Zirconium (Zr)
Yield Strength
Silicon (Si)
Tensile Strength
Magnesium (Mg)
Elongation
Copper (Cu)
Aluminium (Al)
Manganese (Mn)
Iron (Fe)
Titanium (Ti)
Chromium (Cr)
Nickel (Ni)
Vanadium (V)
Silver (Ag)
Boron (B)
Bismuth (Bi)
Cadmium (Cd)
Cobalt (Co)
Erbium (Er)
Europium (Eu)
Gallium (Ga)
Lead (Pb)
Lithium (Li)
Tin (Sn)
Beryllium (Be)
(Other Trace)

(g)

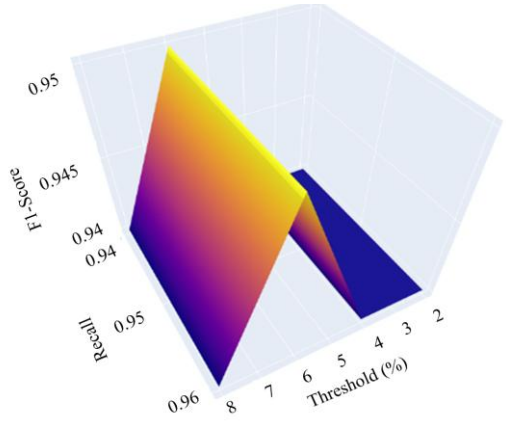


(h)

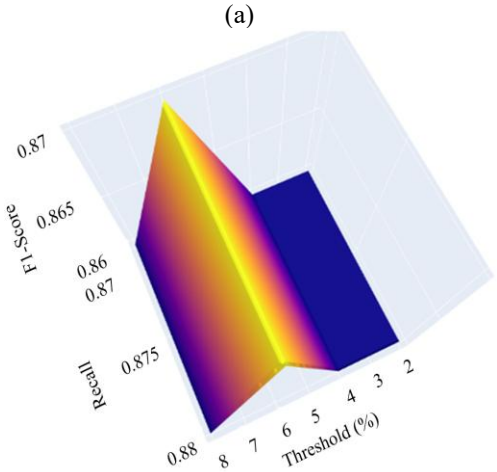
Figure 16. Class-wise feature importance profile (a) No processing, (b) Peak aged, (c) Over aged, (d) Strain hardened, (e) Naturally aged, (f) Strain hardened, (g) Artificial aged, (h) Cold worked.



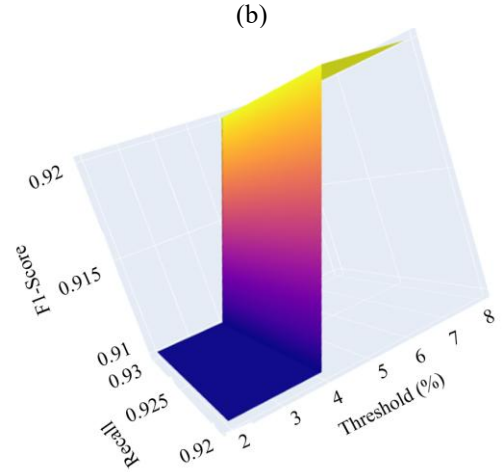
(a)



(b)



(c)



(d)

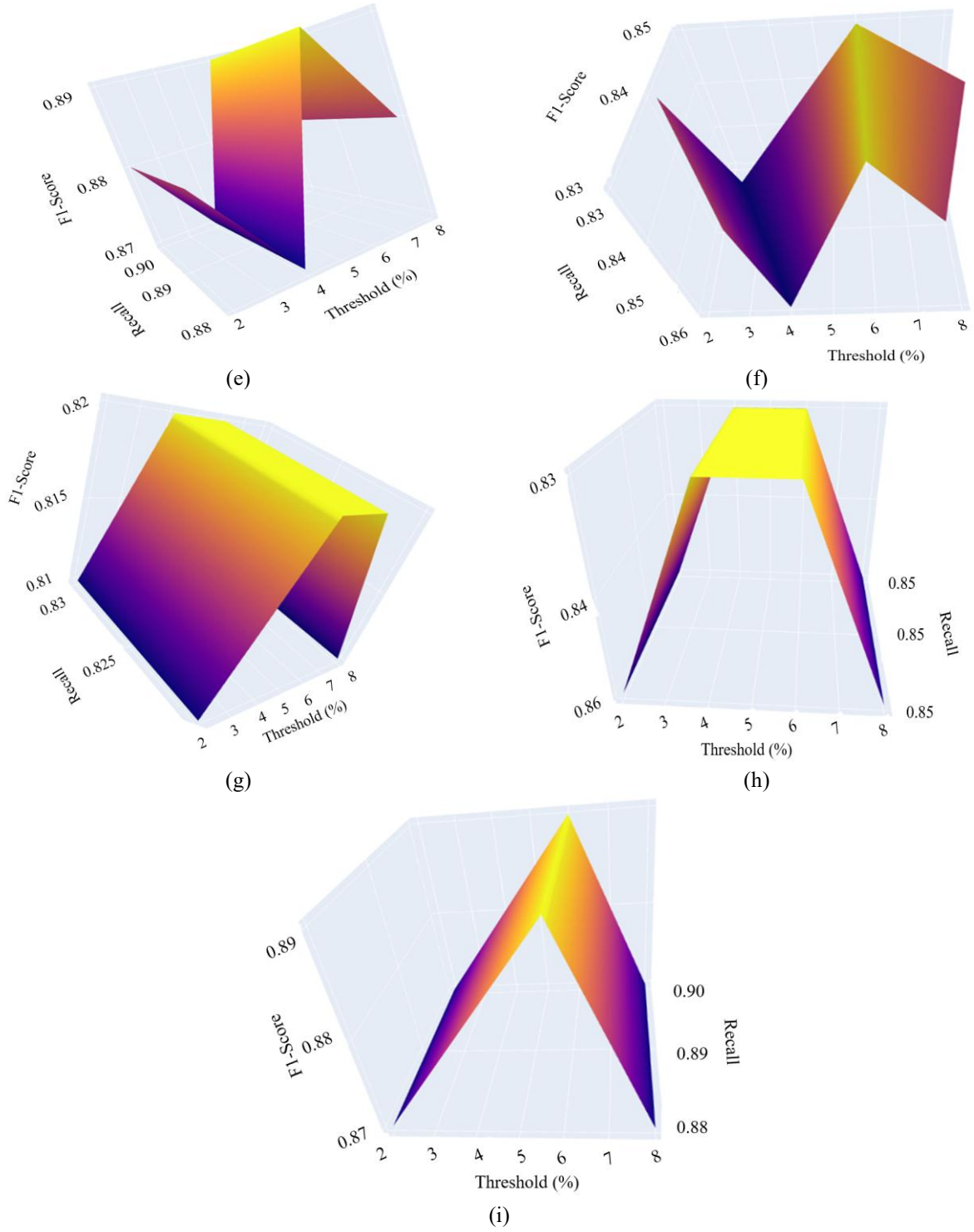


Figure 17. Guided threshold tuning analysis. (a) Class 1, (b) Class 2, (c), Class 3, (d) Class 4, (e) Class 5, (f) Class 6, (g) Class 7, (h) Class 8, (i) Macro average.

5. Conclusion

The problem of the inverse design prediction Al alloy based on post processing from property and compositional data was addressed in this paper. The highly imbalanced and noisy real-world material dataset was utilized in this work. This inverse prediction problem was addressed with manifold-guided VAE-oversampled TabTransformer framework for attention-based and high-performance classification. The 80/20 training-validation split performance results demonstrated that the proposed method not only tackled imbalance issues but also measured significantly accurate predictions. The VAE proposed framework observed a transformative improvement over baseline framework by increasing the macro-average F1-scores from 0.49 to 0.89 and ensured stability by achieving almost balanced weighted F1-scores. This statistically significant ($p < 0.0001$) high performance of this framework was the result of consistently handling both majority and minority classes. Through context-specific metallurgical reasoning, the proposed framework efficiently learned to reconstruct continuous nature of the input from discretized representations. The prediction outcome of the proposed framework was found to be superior to the existing method (Barnard, 2024), where accuracy was varied from 0.60 to 0.94 with dominance from binary classes but limited to non-linear multi-classes only. The proposed framework worked significantly well for multi-class classification with optimum accuracy and macro F1-scores. With non-linear data manifolds, VAE framework generated dense, high fidelity and practical data samples and overcame the linear interpolation limitations of existing methods. The VAE-framework combined with TabTransformer deep learning model transformed a classifier into a stable, well-calibrated, powerful and highly accurate predictive tool suitable for inverse design problems in imbalanced domains. In future, the work can be extended to other material systems with imbalanced datasets. With more advanced algorithms, the work can also be applied to classification of functional properties in polymers, synthesizability prediction in inorganic compounds and anomaly detection in structural steel failure modes.

Conflicts of Interest

The authors confirm that there is no conflict of interest to declare for this publication.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors would like to thank the editor and anonymous reviewers for their comments that help improve the quality of this work.

AI Disclosure

During the preparation of this work the author(s) used generative AI in order to improve the language of the article. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

References

Ai, Q., Wang, P., He, L., Wen, L., Pan, L., & Xu, Z. (2023). Generative oversampling for imbalanced data via majority-guided VAE. In *International Conference on Artificial Intelligence and Statistics* (pp. 3315-3330). PMLR. Valencia, Spain.

Attari, V., & Arroyave, R. (2025). Decoding non-linearity and complexity: deep tabular learning approaches for materials science. *Digital Discovery*, *4*(10), 2765-2780. https://doi.org/10.1039/D5DD00166H.

Aurpa, T.T., Fariha, K.N., Hossain, K., Jeba, S.M., Ahmed, M.S., Adib, Md. R.S., Islam, F., & Akter, F. (2024). Deep transformer-based architecture for the recognition of mathematical equations from real-world math problems. *Heliyon*, *10*(20), e39089. https://doi.org/10.1016/j.heliyon.2024.e39089.

Badaro, G., Saeed, M., & Papotti, P. (2023). Transformers for tabular data representation: a survey of models and applications. *Transactions of the Association for Computational Linguistics*, *11*, 227-249. https://doi.org/10.1162/tacl_a_00544.

Bang, K., Kim, J., Hong, D., Kim, D., & Han, S.S. (2024). Inverse design for materials discovery from the multidimensional electronic density of states. *Journal of Materials Chemistry A*, *12*(10), 6004-6013. https://doi.org/10.1039/D3TA06491C.

Barnard, A.S. (2024). Inverse prediction of Al alloy post-processing conditions using classification with guided oversampling. *Machine Learning: Science and Technology*, *5*(4), 045060. https://doi.org/10.1088/2632-2153/ad95dc.

Bhat, N., Barnard, A.S., & Birbilis, N. (2023a). Improving the prediction of mechanical properties of aluminium alloy using data-driven class-based regression. *Computational Materials Science*, 228, 112270. https://doi.org/10.1016/j.commatsci.2023.112270.

Bhat, N., Barnard, A.S., & Birbilis, N. (2024a). Inverse design of aluminium alloys using multi-targeted regression. *Journal of Materials Science*, *59*(4), 1448-1463. https://doi.org/10.1007/s10853-023-09317-2.

Bhat, N., Barnard, A.S., Birbilis, N (2023b). Aluminium alloy dataset for supervised learning. Mendeley Data, V1, doi: 10.17632/b6br4yk6r3.1. https://data.mendeley.com/datasets/b6br4yk6r3/1.

Bhat, N., Birbilis, N., & Barnard, A.S. (2024b). Unsupervised learning and pattern recognition in alloy design. *Digital Discovery*, *3*(12), 2396-2416. https://doi.org/10.1039/D4DD00282B.

Boyer, R.R. (1995). Titanium for aerospace: rationale and applications. *Advanced Performance Materials*, *2*(4), 349-368. https://doi.org/10.1007/BF00705316.

Brishti, F., Zhang, F., Mohammed, S., Bai, L., Wu, F., & Chen, B. (2025). Imbalanced classification with label noise: a systematic review and comparative analysis. *ICT Express*, *11*(6), 1127-1145. https://doi.org/10.1016/j.icte.2025.09.011.

Butler, K.T., Davies, D.W., Cartwright, H., Isayev, O., & Walsh, A. (2018). Machine learning for molecular and materials science. *Nature*, *559*(7715), 547-555. https://doi.org/10.1038/s41586-018-0337-2.

Chawla, N.V., Bowyer, K.W., Hall, L.O., & Kegelmeyer, W.P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321-357. https://doi.org/10.1613/jair.953.

Gosain, A., & Sardana, S. (2017). Handling class imbalance problem using oversampling techniques: a review. In *2017 International Conference on Advances in Computing, Communications and Informatics* (pp. 79-85). IEEE. Udupi, India. https://doi.org/10.1109/ICACCI.2017.8125820.

Han, X.-Q, Wang, X.-D., Xu, M.-Y., Feng, Z., Yao, B.-W., Guo, P.J., Gao, Z.-F., & Lu, Z.-Y. (2025). AI-driven inverse design of materials: past, present, and future. *Chinese Physics Letters*, *42*(2), 027403. https://doi.org/10.1088/0256-307X/42/2/027403.

Jain, A., Ong, S.P., Hautier, G., Chen, W., Richards, W.D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G., & Persson, K.A. (2013). Commentary: the materials project: a materials genome approach to accelerating materials innovation. *APL Materials*, *1*(1), 011002. https://doi.org/10.1063/1.4812323.

Jha, D., Ward, L., Paul, A., Liao, W., Choudhary, A., Wolverton, C., & Agrawal, A. (2018). ElemNet: deep learning the chemistry of materials from only elemental composition. *Scientific Reports*, *8*(1), 17593. https://doi.org/10.1038/s41598-018-35934-y.

Jiang, J., Zhang, C., Ke, L., Hayes, N., Zhu, Y., Qiu, H., Zhang, B., Zhou, T., & Wei, G.-W. (2025). A review of machine learning methods for imbalanced data challenges in chemistry. *Chemical Science*, *16*(18), 7637-7658. https://doi.org/10.1039/D5SC00270B.

Lee, M. (2023). A mathematical interpretation of autoregressive generative pre-trained transformer and self-supervised learning. *Mathematics*, *11*(11), 2451. https://doi.org/10.3390/math11112451.

Liu, Q., Khalil, M., Jovanovic, J., & Shakya, R. (2024). Scaling while privacy preserving: a comprehensive synthetic tabular data generation and evaluation in learning analytics. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 620-631). ACM, Kyoto, Japan. https://doi.org/10.1145/3636555.3636921.

Lu, S., Zhou, Q., Chen, X., Song, Z., & Wang, J. (2022). Inverse design with deep generative models: next step in materials discovery. *National Science Review*, *9*(8). nwac111. https://doi.org/10.1093/nsr/nwac111.

Madani, M., Lacivita, V., Shin, Y., & Tarakanova, A. (2025). Accelerating materials property prediction via a hybrid transformer graph framework that leverages four body interactions. *Npj Computational Materials*, 11(1), 15. https://doi.org/10.1038/s41524-024-01472-7.

Magdassi, S., Grouchko, M., & Kamyshny, A. (2010). Copper nanoparticles for printed electronics: routes towards achieving oxidation stability. *Materials*, *3*(9), 4626-4638. https://doi.org/10.3390/ma3094626.

Noh, J., Kim, J., Stein, H.S., Sanchez-Lengeling, B., Gregoire, J.M., Aspuru-Guzik, A., & Jung, Y. (2019). Inverse design of solid-state materials via a continuous representation. *Matter*, *1*(5), 1370-1384. https://doi.org/10.1016/j.matt.2019.08.017.

Polmear, I.J. (2005). Light alloys: from traditional alloys to nanocrystals. Elsevier. Oxford, UK. https://doi.org/10.1016/B978-0-7506-6371-7.X5000-2.

Prasad, N.E., & Wanhill, R.J.H. (2017). *Aerospace materials and material technologies*. Springer, Singapore. https://doi.org/10.1007/978-981-10-2134-3.

Ramprasad, R., Batra, R., Pilania, G., Mannodi-Kanakkithodi, A., & Kim, C. (2017). Machine learning in materials informatics: recent applications and prospects. *Npj Computational Materials*, *3*(1), 54. https://doi.org/10.1038/s41524-017-0056-5.

Sharma, H., & Gosain, A. (2023). Oversampling methods to handle the class imbalance problem: a review. In *International Conference on Soft Computing and its Engineering Applications* (pp. 96-110). Springer Nature, Switzerland. https://doi.org/10.1007/978-3-031-27609-5_8.

Sharma, S., Gosain, A., & Jain, S. (2022). A review of the oversampling techniques in class imbalance problem. In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021* (Vol. 1, pp. 459-472). Springer, Singapore. https://doi.org/10.1007/978-981-16-2594-7_38.

Tang, X., Li, W., Sun, W., Wen, T., & Guo, K. (2025). Unbalanced data oversampling method based on improved VAE-CGAN. In *2025 5th International Symposium on Computer Technology and Information Science* (pp. 200-208). IEEE. Xi'an, China. https://doi.org/10.1109/ISCTIS65944.2025.11066045.

Wahid, M.A., Siddiquee, A.N., & Khan, Z.A. (2020). Aluminum alloys in marine construction: characteristics, application, and problems from a fabrication viewpoint. *Marine Systems & Ocean Technology*, *15*(1), 70-80. https://doi.org/10.1007/s40868-019-00069-w.

Wang, A.X., & Nguyen, B.P. (2025). TTVAE: transformer-based generative modeling for tabular data generation. *Artificial Intelligence*, *340*, 104292. https://doi.org/10.1016/j.artint.2025.104292.

Wang, J., Wang, Y., & Chen, Y. (2022). Inverse design of materials by machine learning. *Materials*, *15*(5), 1811. https://doi.org/10.3390/ma15051811.

Wang, X., Zhang, S., Jiang, D., Yu, W., Zheng, Y., Luo, C., Wang, H., & Wang, Z. (2025a). Transformer-based multimodal learning for predicting mechanical properties in heat-treated stainless steel. *Materials & Design*, *259*, 114800. https://doi.org/10.1016/j.matdes.2025.114800.

Wang, Z., Zhu, J., Ma, T., Huang, X., & Wang, S. (2025b). SDR-TabTransformer: a novel deep learning algorithm for mapping of roadway area surface deformation combined with SBAS-InSAR. *International Journal of Pavement Engineering*, *26*(1), 2539493. https://doi.org/10.1080/10298436.2025.2539493.

Zhang, H. (2011). *Building materials in civil engineering*. Woodhead Publishing, New Delhi. https://doi.org/10.1533/9781845699567.206.

Zhuang, Z., & Barnard, A.S. (2023). Structure-free mendeleev encodings of material compounds for machine learning. *Chemistry of Materials*, *35*(21), 9325-9338. https://doi.org/10.1021/acs.chemmater.3c02134.

Zhuang, Z., & Barnard, A.S. (2024). Classification of battery compounds using structure-free Mendeleev encodings. *Journal of Cheminformatics*, *16*(1), 47. https://doi.org/10.1186/s13321-024-00836-x.

Zivic, F., Malisic, A.K., Grujovic, N., Stojanovic, B., & Ivanovic, M. (2025). Materials informatics: a review of AI and machine learning tools, platforms, data repositories, and applications to architectured porous materials. *Materials Today Communications*, *48*, 113525. https://doi.org/10.1016/j.mtcomm.2025.113525.