ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

A Novel Approach to Dairy Sales Forecasting: Multi-Perspective Fusion Bi-LSTM Coupled with Universal Scale CNN

Naveen D. Chandavarkar[1] and Soumya S[2]

ABSTRACT

In today's world, accurately predicting sales is important for minimising costs and improving overall profits. A wide range of understanding of organisational performance and future sales can be accomplished through improved customer service strategies. It aids in enhancing product returns and lowering lost sales, leading to more efficient production planning. The sales prediction in dairy products reflects distinctive challenges, predominantly due to the quality of these products, which is closely connected to consumers' health. To overcome the problem, the proposed research employs an effective DL (Deep Learning) based technique for forecasting the sales of dairy products by analysing the dairy goods sales dataset from an openly available website. The proposed research utilises Universal Scale CNN (Convolutional Neural Network), a 1D CNN, which is capable of learning the features at optimal and effective rates. The following features are passed to the Multi-Perspective based Bi-LSTM (Bidirectional Long Short-Term Memory), which is capable of learning features effectively by reducing error rates in predicting sales rates of dairy-based products. The overall performance of the proposed Multi-Perspective Fusion Bi-LSTM with Universal Scale CNN is evaluated with performance metrics, including RMSE (Root Mean Squared Error), MAE (Mean Absolute Error) and MSE (Mean Squared Error). These performance metrics evaluate the proposed model's effectiveness in forecasting dairy product sales.

Article information:

Keywords: Dairy Sales Forecasting, DL (Deep Learning), Universal Scale CNN (Convolutional Neural Network), 1D CNN, Bi-LSTM (Bidirectional Long Short-Term Memory)

Article history:
Received: May 30, 2025
Revised: August 22, 2025
Accepted: October 10, 2025
Published: October 31, 2025 (Online)

DOI: 10.37936/ecti-cit.2025194.262337

1. INTRODUCTION

Recently, the dairy industry has been experiencing some intense changes all over the world. According to the information given by the dairy price index of the United Nations Food and Agriculture Organization, the current cost has been lowered by up to 26% to others, in the month of February 2014 [1]. The milk prices of China have plunged recently, due to trade constraints imposed against Russia in the market, using some milk products. Due to this, the powdered milk production is on the upsurge and by 2025, it will have achieved 177 million metric tons with an 18% increase per year. In emerging countries, growing urban populations and decent living standards could be driving this growth [2]. The European farmers have started employing intervention stocks in an effort to hedge against the universal deterioration in milk prices. India is the paramount milk producer which producing 23% of the world's milk. In the next few years, the national milk production growth rate is forecasted to be almost 6.2%. The highest milk production state of India is UP (Uttar Pradesh) with nearly 18% [3]. The overall quantity of milk produced by India is from UP. The most milk-producing states in India are UP, AP (Andhra Pradesh), MP (Madhya Pradesh), Rajasthan and Gujarat [4]. The division of Animal Husbandry and Dairy proclaimed finance in infrastructure development in June 2020 as

[1,2]The authors are with the Institute of Computer Science, Srinivas University, Mangalore, India, Email: Naveendchandavarkar.cet@srinivasuniversity.edu.in and soumyas.ccis@srinivasuniversity.edu.in
[1]The author is with the Government First Grade College, Lingasugur, Raichur District, India, Email: Naveendchandavarkar.cet@srinivasuniversity.edu.in
[1]Corresponding author: Naveendchandavarkar.cet@srinivasuniversity.edu.in

a representative of the Indian government. In order to predict milk production, fat and protein content, researchers are enhancing scheduling and distribution strategies more precisely and feed intake [5]. Correspondingly, small businesses, programmers and private businesses will become distracted to motivate investment in the animal and dairy feed industries, which is expected to outcome in the creation of 3.5 million new jobs. Researchers see Indian Food manufacturing initiatives emerging to the point where researchers can strive with the supreme in the globe, and Indian food brands are flourishing increasingly eminent on international markets with these funds [6]. There is a firm requirement for innovative approaches for milk process enhancement in the dairy industry. The creators of dairy products are required to direct their procedures with utmost caution due to the milk components of milk being reduced rapidly [7].

The AI (Artificial Intelligence) is the best method to tackle this big data-relevant problem [8]. Huge datasets are used by recognized ML algorithms to instinctively analyse and discover patterns in variables. The ML techniques can analyse existing unnoticed patterns in data, generate an exclusive understanding and direct researchers in the right direction [9]. The concept of AI, known as ML, uses complex algorithmic frameworks to resolve issues that could otherwise be unsolvable by computers. A huge retailer could have thousands of distributors, as opposed to the lesser and far more controllable number of producers. The supply chain of retail could be able to activated effectively as a result, and it should highlight the coordination of the supply chain more [10]. Additionally, a retailer competes with millions of end users against a few distributors. Retailers are required to spend more time appropriately planning and responding to understand customer demand. There are variations among the cost structures of retailers and manufacturers. Among several fields, the dairy sector is the one where ML could be employed. The sequence of processes that take place during the time period raw milk is obtained and kept in the supermarket is known as the dairy supply chain [11]. The most significant factor that impacts the cost of commercial dairy products is the price variation of raw materials, which impacts the price of dairy products to a certain extent. There are various time series methods to predict the price, and one of the conventional regression methods is Ridge, AR, Lasso, ARIMA (Autoregressive Integrated Moving Average), etc [12]. The ARCH (Autoregressive Conditional Heteroskedasticity) is a statistical model for time series data that determines the variance of innovation or current error term as a function of the definite sizes of the error of the existing time period. With the development of AI technology, the time series prediction techniques depicted by the BP NN (Neural Network), LSTM (Long Short-Term Memory), SVM (Support Vector Machine), RNN (Recurrent NN) and transformers have been broadly employed [13]. The DL techniques can analyse the pattern and structure of data for the complexity and nonlinearity of time series prediction. Likewise, the LSTM can identify and analyse the patterns or interactions in data by a self-learning process; however, the computation time will be huge, and the time consumption of the approached network is deep [14]. Though various unsupervised time series approaches are similar to NLP (Natural Language Processing). CV (Cross-validation) fields, which have robust inductive bias they are not appropriate for modelling time series in most cases [15]. To tackle this problem, the proposed model employs 1D–CNN (Convolutional NN) and Bi-LSTM (Bidirectional LSTM), which are enhanced with Multi-Perspective Fusion Bi-LSTM with Universal Scale CNN to predict the sales of dairy products. The proposed research employs a dairy goods sales dataset from an openly available website to evaluate the proposed Multi-Perspective Fusion Bi-LSTM with Universal Scale CNN model. The demonstration of the proposed model has been examined with performance metrics such as RMSE, MAE and MSE.

1.1 Research Contribution

The main objective of the proposed research is as follows:

- To use the dairy goods scales dataset for forecasting the sales of the dairy products.
- To employ different pre-processing techniques like feature scaling, label encoding and outlier removal for handling missing values, removing inconsistencies of the data.
- To implement an improved regression model, such as proposed multi-perspective fusion BiLSTM with Universal Scale CNN model for precisely predicting the sales of dairy products with the least error rates.
- To exhibit the proposed Universal Scale CNN model to track and conceal all the scales of the respective field by converting the input among innumerable combinations of the prime-sized kernels.
- To evaluate the performance-based metrics of the model performance enabling regression for the sales forecast of dairy products.

1.2 Paper Organization

The following paper is divided into four sections. Section II deliberates the existing approaches for forecasting the sales of dairy. Section III deliberates on the process of the proposed method. Section IV reflects the results and discussion of the proposed model. Section V reflects the conclusion and future work of the proposed model.

2. LITERATURE REVIEW

The existing study [16] has utilized several algorithms, namely the RF (Random Forest), GB (Gradient Boosting), SVM, NN, and AdaBoost, for forecasting the price of milk. Additionally, the ML approach has included the HPT (Hyper Parameter Tuning) with nested cross-validation for estimating the validity of the suggested algorithms. The approached algorithms have attained better outcomes in contrast with the others. Likewise, the study [17] has suggested an RMP-CPR (Raw Milk Price Prediction Framework), which has been utilized for examining the performance of the consumer regarding the relationship between milk price and dairy consumption. As a result, the RMP-CPR has enhanced the development of accurate marketing strategies. Furthermore, the framework has played a major role in balancing the interests among consumers, dairy enterprises, and dairy farmers. Similarly, the study [18] has implied that NN classification is the right direction for forecasting. As ANN (Artificial Neural Networks) has been influenced by the significance and transparency of the trained input data, the study has suggested a methodology named CS (Compressed Sensing), which has considered the incomplete data as noisy trends that are to be reconstructed using CS reconstruction algorithms. Finally, the suggested methodology has given effective outcomes. Likewise, the study [19] has introduced an ML technique for analysing the PPI (Producer Price Index) of cheese manufacturing companies. Several DL models, such as LSTM, BILSTM, and GRU, have been approached and tested. Furthermore, in contrast with the other performance metrics, the LSTM model has performed better than the others. As a result, the suggested model is similar to the BILSTM and GRU models, which have delivered better outcomes using the PPI. The LSTM has been implemented by the study [20] for predicting the cow cheese production by using ML and DL techniques, which have not been implemented in the field previously. Along with LSTM, a few models like MLP, SVR and KNN have been tested and compared with performance metrics such as MSE, MAE, MAPE (Mean Absolute Percentage Error), and RMSE. As a result, the LSTM model has demonstrated high accuracy and dependability in the production of cow cheese. Similarly, the study [21] has implemented the ARIMA model for forecasting the milk price in the Ukrainian market of raw milk. Additionally, it has demonstrated the absence of a single direction during the bias period. The delivered results of the milk price have been modified regarding the time lag. Finally, the resultant data has been compared with real prices of milk sales, which has specified the presence of irrelevant changes by demonstrating the capability of the suggested approach. Furthermore, the drivers of regional milk prices in Russia have faced numerous challenges in the accessibility of data. Hence, the study [22] has suggested an RF approach to overcome the data-based difficulties in which the traditional panel regression nodes have been limited. Finally, the model training and hyper parameter optimization have been implemented on the trained data set with cross-validation. As a result, the suggested algorithm has a better demonstration in terms of test data.

Correspondingly, the study [23] has performed the modifications in the milk cost and other yields at the European Union (EU), where the descriptive statistics, correlation and regression analyses have been presented. Additionally, the approach has utilized the ADF (Dickey-Fuller) test and the GARCH (Generalised Auto Regressive Conditional Heteroscedasticity) model, which have been used to evaluate the normality and fluctuations of the data. As a result, the processes for butter, WMP (Whole Milk Powder), SMP (Skim Milk Powder), cheddar, Edam, and Gouda have been reliant on earlier values. Furthermore, the main goal of the study [24] has implied that the milk supply from Norwegian dairy farmers to dairy has been performed with the utilisation of a time series model in an ML approach. According to the results, the application forecast monthly milk deliveries to dairies up to 24 months in advance with a MAE of 1-2 per cent. Additionally, the outcomes have delivered useful insights into the characteristics, which have been essential for forecasting future milk delivery. Similarly, the existing study [25] has examined the cost of cheddar cheese, butter, SMP, and WMP with the utilisation of an MFSV (Multivariate Factor Stochastic Volatility) technique for extracting correlation matrices and time-varying covariance. As a result, the matrices delivered a higher reliance during the mid-year of 2006-2014, which has been recognised by the RADP (Regional Agricultural Dairy Policies). Similarly, the study [26] has evaluated the cost volatility of four dairy commodities, MP, butter, WMP and cheddar cheese, with the utilisation of a panel-GARCH model for the first time. Furthermore, it has delivered potential efficiency received from assessing conditional variances and covariance's. As a result, the CV has increased values during 2007-2016, when conditional cross-correlations have been low. Furthermore, on the basis of examining the performance of volatility spill over and the rate of import trade, GDT milk powder has been designated. The study [27] has suggested the GARCH 1:1 model, namely BEKK (Baba-Engle-Kraft-Kroner), for implementing GDT milk powder. As a result, a bidirectional volatility of spill over has been observed between China's raw milk market and GDT milk powder. Likewise, the study [28] has derived the certified data from Russia from 2015-2019, which has been examined by 12 predictor variables for describing the milk price of the particular region. Also, with the utilization of the ML algorithm, the MT and HO (Hyper-parameter Optimization) have been accomplished through SCV (Spatiotemporal Cross-Validation) technique. As a result, the suggested RF algorithm has delivered better outcomes by exposing the four main factors, such as population density, production of milk, income and livestock numbers that have caused dissimilarity in Russia's raw milk cost. The study [29] has suggested LSTM-based NNs for improving the decision-making process. The training data has been obtained from more than 6000 different herds. As a result, the approach has performed better than the ARIMA statistical model by 68% at verification time, forecasting the monthly income for each cow's fifth year with an RMSE of 8.36.

The study [30] has suggested a SAE (Sequential Auto Encoder) for decoding low-dimensional representations and sequential data. Furthermore, the SAE, in contrast with the conventional MLP (Multi-Layer Perceptron) model, has utilized parity data and protected milk yields as input. As a result, the combination of lactation number, herd statistics, daily milk yields, as well as reproduction and health events the cow has faced in the lactation period has been displayed in qualitative latent representations. The existing study [31] has demonstrated the designs and determinants of organic dairy payments and evaluated the monthly cost premiums for organic milk, eggs, fluid milk, and yoghurt, using ML techniques over the period 2008–2017. As a result, the calculated premiums are 27%, 56%, 38% and 51%, respectively, along with a CAGR (Compound Annual Growth Rate) of 8%, 7%, −1.42%, and −2.86%. Furthermore, keeping organic sales constant, the premium results are high. The study [32] has explored an approach for forecasting milk yield and lactation patterns in different phases, utilising a dataset from the largest dairy farms in Jordan. The data has been reinforced in the development of data-driven AI and ML models. As a result, the approached models had better outcomes in increasing the dairy cattle efficiency. The study [33] has implied that the TMC (Tamil Nadu Milk Cooperatives) has made numerous donations to the development of the dairy sector. Additionally, the study has forecasted milk production in Tamil Nadu using TSM by utilizing annual milk data from 1976 to 2020, and also implemented ARIMA and ANN for forecasting milk production. As a result, the CAGR of the predicted milk production is 0.02% with the utilization of performance metrics such as RMCE, MAE, and MAPE. The study [34] has suggested an Edge-Centric approach, App Adapt-LWDL (Application-Adaptive Light-Weight DL) for identifying the type of milk and recognising adulteration. Moreover, the approached algorithm has maintained speed and accuracy by influencing the clipping ratio. As a result, the suggested framework has enhanced efficiency and accuracy regarding the application.

2.1 Problem Identification

- The limitation of the existing approach [18] implies that the advanced compilation of the data and retrieving characteristics has to be accomplished using a real-world COVID-19 data set.
- In contrast, GRU models did not adopt the LSTM and BILSTM frameworks [19].
- The limitation of the prevailing study [29] implies that more detailed data and incorporation of raw variables have to be included.

3. RESEARCH METHODOLOGY

3.1 Proposed Method

Sales forecasting is a critical task across various industries. Accurate forecasting helps alleviate inventory and stock-related challenges, thereby preventing revenue loss. Hence, the proposed model utilizes Multi-Perspective Fusion Bi-LSTM with Universal Scale CNN to effectively forecast dairy sales.

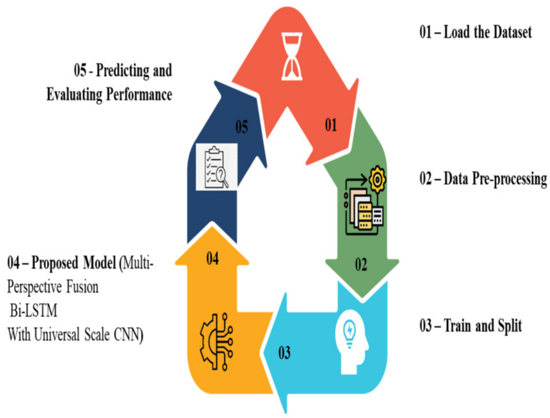

Fig.1: *Overall Flow of Proposed Research.*

Figure 1 depicts the flow of the proposed research. Initially, the dairy sales dataset is loaded as input into the proposed model. The loaded dataset undergoes several pre-processing techniques such as feature scaling, label encoding and outlier removal, which enhance model performance by transforming raw data into a structured format that improves the models efficiency, reliability and accuracy. Next, the processed dataset is divided into two subsets: training data (80%) and testing data (20%). The training data is processed using Multi-Perspective Fusion Bi-LSTM with Universal Scale CNN, while the testing data is used for the evaluation of the proposed model. Finally, the performance of the proposed model is assessed using performance metrics such as RMSE, MAE and MSE.

3.2 Dataset Description

The dataset used in this research is sourced from the open-access website Kaggle. The link of the dataset is provided below for reference: `https://`

www.kaggle.com/datasets/suraj520/dairy-goods-sales-dataset/data. This dataset offers a comprehensive and detailed collection of data relevant to inventory management, dairy farms, dairy sales and products. It encompasses a wide range of information including farm size, farm location, cow population, land area, product details, production dates, storage conditions, brand information, pricing, quantities, sales information, expiration dates, sales channels, customer locations, reorder qualities, stock qualities and stock thresholds.

3.3 Data Pre-processing

3.3.1 Feature Scaling

This technique is significant when the dataset encompasses features with varying scales or units. It assures that all features are subsidized for analysis using z-score normalization and min-max scaling. The min-max scaling technique is used to adjust feature values to a fixed range, typically [0, 1]. Meanwhile, the standardization technique transforms the data to have mean of 0 and SD (Standard Deviation) of 1. These two methods helps in normalizing the data, which can improve the performance of the proposed model.

3.3.2 Label Encoding

Label encoding is utilized for categorical variables that need to be transformed into a numerical format for the proposed algorithms. This process involves assigning integer values to each category using specific libraries. Each unique entry in a categorical feature is assigned a distinct integer, which means milk, yogurt and cheese are encoded as 0, 1 and 2. The python library sklearn provides "Label Encoder" that automates the transformation of categorical variables into a suitable format.

3.3.3 Outlier Removal

This technique can significantly distort results and affect the performance of the proposed model. Analysing and removing outliers from the dataset involves the use of statistical methods and visualization techniques. The statistical method applies standard techniques such as z-score and IQR (Interquartile Range). Values with a z-score less than -3 or greater than 3 are identified as outliers. Visualization techniques, such as scatter plots or box plots can help visualize the outliers before making decisions regarding their removal.

3.4 Multi-Perspective Fusion Bi-LSTM with Universal Scale CNN

To forecast the dairy sales, the proposed research utilizes a Multi-Perspective Fusion Bi-LSTM with Universal Scale CNN. The complete process of the proposed model is represented in Figure 2.

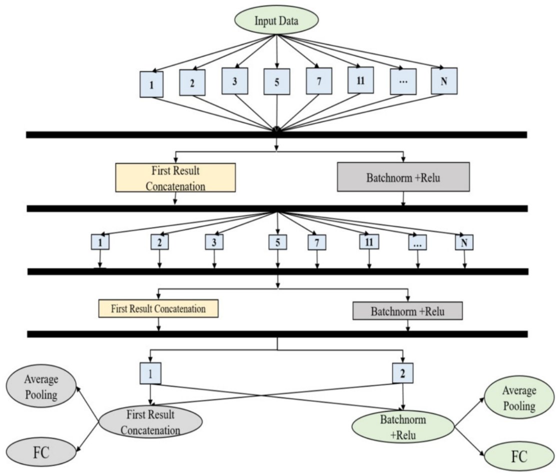

Fig.2: *The Entire Process of Proposed Model.*

Figure 2 illustrates the complete process of the proposed model. The model begins with an input data layer structured into sequential blocks that likely represent a series of data values or features. The data series is divided into two sections. The first section involves the process of first result concatenation, where the initial data blocks are combined. The output is then processes through a batch normalization layer followed by a ReLU activation function layer. Batch normalization stabilizes the training process by normalizing the data, while the ReLU introduces non-linearity. The second section performs a similar operation, combining the initial data blocks and processing the output through batch normalization and ReLU. This signifies a parallel operation to capture multiple aspects of the data.

After the batch normalization and ReLU activation function in both sections, the outputs are concatenated again. This layer-by-layer approach with transitional concatenations suggests a deep structure model, where each level gathers features from previous layers to capture more complex patterns. Following the initial transformations, both sections undergo average pooling, which reduces dimensionality by averaging values, thus retaining significant features while discarding less critical information. Followed by the pooling process, each section connects to a fully connected layer, usually in classification tasks. This layer combines all the processed features to deliver the final classifications or predictions. The provided a time seriesIn, smearing a neural convolution process with kernel γ is corresponding smearing a signal processing convolution with kernel ψ. Besides, in the Fourier domain this process is also corresponding to the element-wise multiplication among In and ψ.

$$In * \gamma = In \otimes \phi = F^{-1}(F(In) \cdot F(\psi)) \tag{1}$$

Where, the signal processing convolution of In and ψ is,

$$(In \otimes \psi)[P] = \sum_{m=0}^{M-1} In[P-Q]\psi[Q] \tag{2}$$

Where, the signal processing convolution is denoted as $\otimes$, the convolutional theorem is stated as

$$In \otimes \psi = F^{-1}(F(In) \cdot F(\psi)) \tag{3}$$

Where, the Fourier transform is represents as F and the inverse Fourier transform is represents as F^{-1}. In a traditional 1D-CNN, the convolution of In and γ is,

$$(ln * \gamma)[P] = \sum_{m=0}^{M-1} In[P-Q]\gamma[R-1-Q] \tag{4}$$

Where, a neural convolution is denoted as *. When $\psi[Q] = \gamma[R-1-Q]$ in equation 1 which is affirmed in the Lemma 1. $\psi[Q] = \gamma[R-1-Q]$ states ψ has the reversed sequence of γ. Which means $\gamma = [1,2,3]$ where as $\psi = [3,2,1]$. In the frequency domain, the proposed research analyze the time series, which can decompose a time series data In into three divisions.

$$F(In) = F(In_s) + F(In_{N_s}) + F(ln_{N_o}) \tag{5}$$

Where, the preferred signal for classification task is denoted as In_s, the noise on a similar frequency of the preferred signal is denoted as In_{N_s} and the noise on the other frequencies is denoted as In_{N_o}. Similarly, the neural convolution kernel γ can be decomposed as.

$$F(\psi) = F(\psi_S) + F(\psi_{N_s}) + F(\psi_{N_o}) \tag{6}$$

To streamline, replace $F(In_s), F(In_{N_s}), F(In_{N_o})$ with A, B, C. Likewise replace $F(\psi_s), F(\psi_{N_s}), F(\psi_{N_o})$ with a, b, c. Hence in the frequency domain, the result of convolution is.

$$\begin{aligned} F(In * \gamma) = F(In) \cdot F(\psi) = Su + Sv + Sc + \\ Tu + Tv + Tc + Ku + Kv + Kh \end{aligned} \tag{7}$$

Here, $F(In_{N_o})$ does not intersect with $F(\psi_s)$ in the Fourier domain, hence that:

$$F(In_{N_o}) \cdot F(\psi_s) == 0 \tag{8}$$

For a similar reason in equation (8), the Sc, Tc, Ku, Kv turned as zero, hence, the result of convolution result might be represents as

$$\begin{aligned} Su + Sv + 0 + F(In * \gamma) = Tu + Tv + 0 + \\ 0 + 0 + Kh = (S+T)(u+v) + Kh \end{aligned} \tag{9}$$

Consequently, if the noisy part of Kh is remove or reduce, then the result of convolution operation $F(In * \gamma)$ will have minimal noise that might be advantageous for the classification. Specifically, the entire equation of the layer output before the activation layer:

$$LayerOutput = BN(In * \gamma + bias) \tag{10}$$

Where, the batch normalization is denoted as BN. If the BN is expanded then the equation (10) will be,

$$r(In * \gamma + bias - E) + \beta \tag{11}$$

Where, the parameters of batch normalization is denoted by r and β, in a batch, the mean value of $In * \gamma$ is denoted as E. All the trainable parameters are denoted as $r, bias, E$ and β. Hence equation (11) has a similar analytic form:

$$LayerOutput = r(In * \gamma + \epsilon)$$

Where, the representation of zoom rate is denoted as $r, \epsilon = bias - E + (\beta/r)$ is inductive bias which is assessed using numeric variables which do not contain any information related to frequency. The proposed research maps the layer output into frequency information.

$$F(LayerOutput) = r[F(In * \gamma) + F(\epsilon)] \tag{12}$$

Include equation (9) into equation (12), the result is:

$$F(LayerOutput) = r[(S+T)(u+v) + Kh + F(\epsilon)]$$

ϵ is a real value vector of constant value from the bias and batch normalization definition. The proposed research represents the constant value as $Total\_bias$, hence:

$$F(\epsilon)[P] = Total\_biasp = 00 \quad P \neq 0$$

Where, $F(\epsilon)$ is a real value vector while Kh is a complex value vector, the Kh noise cannot be represented by ϵ in convolution. The sizes of kernel are a prime number from 1 to P in the first two convolutional layers and all kernel sizes are dissimilar to each other in the layer. It is the last layer to achieve a convolution operation in the 3rd convolution layer. With this circumstances, the receptive fields of proposed Universal Scale CNN can shield all applicable integer in $(0, 2P)$. If the stride is 1, the size of receptive field is assessed as,

$$R_{size} = 1 - L + \sum_{l=1}^{L} R_l \tag{13}$$

Where, the total number of convolution layers is denoted as L and the lth layer's kernel size is denoted as R_l. In divergence to traditional CNN, the proposed research employs a global average pooling afterwards multiple convolution layers relatively than include one pooling layer next to each convolution layer. By global average pooling in proposed 1D-CNN which permits the manipulation of the CAM (Class Activation Map) to identify the contribution of data region to the particular labels. The CAM is assessed as:

$$C_i = \frac{1}{P} \sum\nolimits_{P=1}^{R} \ F_{i,P} \tag{14}$$

Where, the output value of ith channel is denoted as C_i, the length of time series is denoted as P and in the ith channel, the nth value is denoted as $F_{i,P}$. According to the extracted features, the final layer of proposed 1D-CNN is a fully connected layer that is the classifier identifying labels. The proposed research develop an ensemble version of 1D-CNN with M classifier. The proposed ensemble method is weighted by the classification probability based on majority voting. Provided input In, the forecasted output c is assessed as,

$$c = \arg\max_{c} \frac{1}{R} \sum\nolimits_{Q=1}^{R} P(\theta_m, In) * I[c = F(\theta_m, In)], \tag{15}$$

Here, the model parameters of the mth classifier is depicted as θ_m, the probability of the forecasted class c is represented as $P(\theta_m, In)$ to that In belongs and I represents the Boolean indicator. The proposed research van minimize the size of the model through substituting kernels with receptive fields and the representation ability of the proposed model is also avoided. If there is no loss of representation ability, $\forall In$ and θ_B, $\exists \theta_A$ makes $A(In, \theta_A) = B(In, \theta_B)$. The input data is denoted as In, the model's parameter is denoted as θ_*. The output result of the model with parameter is denoted as $*(In, \theta_*)$. The proposed research found when $L(C) = L(C_1)+L(C_2)-1$, $C(K) = 1, C(K_1) = \min(K_1, K_2)$, for $\forall In$,K,$\exists K_1$,K_2 makes

$$Conv(In, K) = Cove(Conv(In, K_1), K_2)$$

Where, the input signal is denoted as In, the kernel of the one layer network is denoted as K, the 1^{st} and 2^{nd} layer of the two layer network is denoted as K_1 and K_2, the kernel size of K is denoted as $L(K)$, the output Channel number of K is denoted as $C(K)$ and the convolution results of In and K with batch normalization and ReLU activation is denoted as $Conv(In, K)$. The final multimodal characteristics achieved could unavoidably comprise a large quantity of terminated information, if the interaction of modality information was not deliberated, which could be inappropriate to the prediction task. As a result, the proposed study employs Cross-perceptual with the multi-modal fusion upon Bi-LSTM, the fusion function. The feature matrices for cross-perceptual is assessed to achieve the information matrix. The attention distribution is assessed by line- by-line soft max function.

4. RESULTS AND DISCUSSION

This section deliberates the performance metrics, EDA (Exploratory Data Analysis) and performance analysis on the proposed research. The performance on the proposed model has been assessed using RMSE, MAE and MSE.

4.1 Performance Metrics

4.1.1 RMSE

Root Mean Squared Error (RMSE) is the standard deviation on the differences between the actual and predicted values. Lower RMSE values indicate a better fit on the model to the data, whereas higher RMSE values suggest less accurate predictions and larger errors. RMSE has been calculated using equation (16)

$$RMSE = \sqrt{\sum\nolimits_{i=1}^{N} (acutal - predicted)^2 / N} \tag{16}$$

4.1.2 MAE

MAE aids as an efficient metric for evaluating the accuracy on proposed model accuracy, providing a balance among effectiveness, and simplicity in performance evaluation.

$$MAE = \frac{1}{n} \sum\nolimits_{i=1}^{n} |y_i - \hat{y}_i| \tag{17}$$

4.1.3 MSE

It is the measurement on image excellence metric. If the standards are nearer to zero, the metric dimension has better quality. The formula for MSE has been calculated in equation (18)

$$MSE = \frac{1}{n} \sum\nolimits_{i=1}^{n} (Y_i - Y_i)^2 \tag{18}$$

4.2 EDA (Exploratory Data Analysis)

Figure 3 depicts the total sales on dairy products using product name. The plot displays twelve different types on dairy products, including butter, ghee, curd, yogurt, ice cream, paneer, milk, lassi, cheese, and buttermilk. Butter achieves the highest sales, with 6000 liters/kg, while lassi, cheese, and buttermilk have the lowest sales, each at 4000 liters/kg.

Figure 4 depicts the customer preference using location and sales channels. The plot shows 15 locations and their sales channel. There are three types on sales channels, online retail, and wholesale. The highest revenue is completed using Chandigarh with

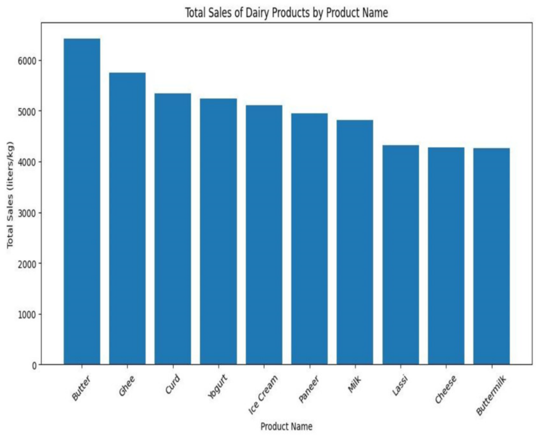

Fig.3: *Total Sales on Dairy Products using Product Name.*

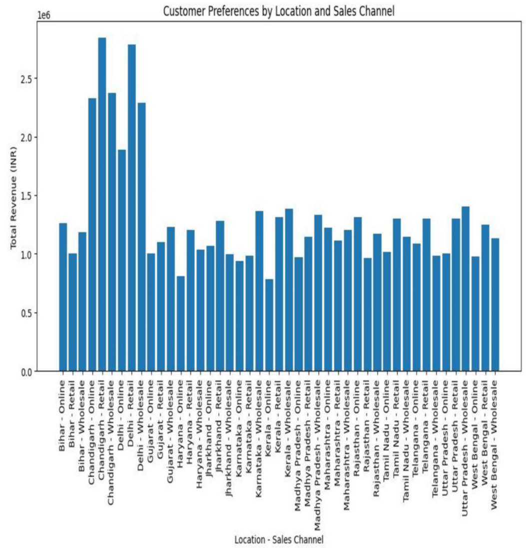

Fig.4: *Customer Preference using Location and Sales Channel.*

2.5 INR in retail sales channel. The lowest revenue is completed using Haryana and Kerala with 0.7 INR in online sales channel.

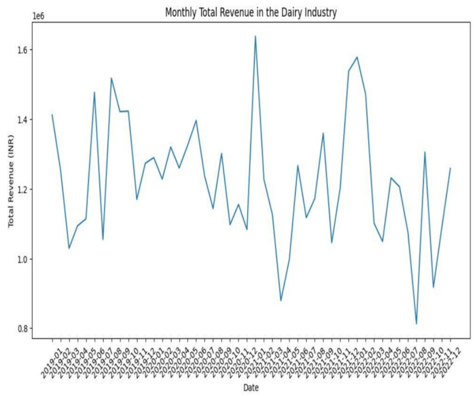

Fig.5: *Monthly Total Revenue in the Dairy Industry.*

Figure 5 depicts the monthly total revenue in the dairy industry. The chart shows total revenue from January 2019 to December 2022.

4.3 Performance Analysis

Table 1: *Performance Analysis on Proposed Model.*

Model	MAE	MSE	RMSE	R-squared
Proposed Model	0.0908	0.016	0.1264	0.9836
RF	0.1513	0.0389	0.1973	0.9601
Linear Regression	0.2102	0.1003	0.3167	0.897
Support Vector Regression	0.2075	0.0989	0.3144	0.8986
Decision Tree	**0.1347**	**0.0371**	**0.1926**	**0.9619**

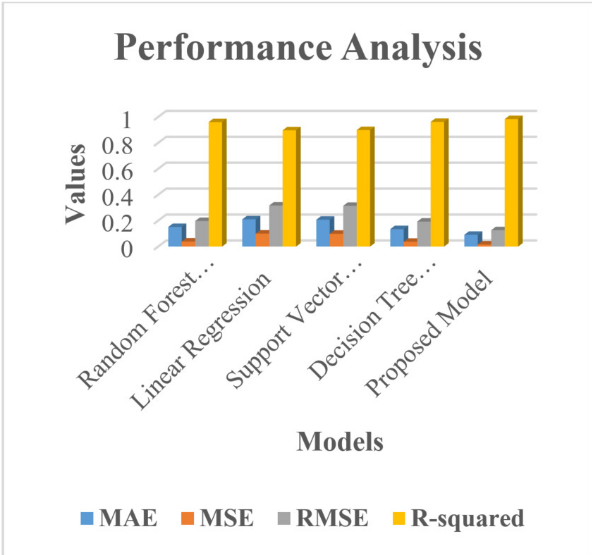

Fig.6: *Performance Analysis on Proposed Model.*

Table 1 represents the performance analysis on the proposed model. The performance on the proposed Multi-Perspective Fusion Bi-LSTM with Universal Scale CNN has been compared with other regression models namely RF Regressor, linear regression, SVR (Support Vector Regression) and DT Regressor (Decision Tree Regressor) with the metrics on MAE, MSE, RMSE, and R-Squared. The values on RFR is 0.1513, 0.0389, 0.1973, and 0.9601. The value on linear regression is 0.2102, 0.1003, 0.3167, and 0.897. The values on SVR are 0.2075, 0.0989, 0.3144, and 0.8986. The values on DTR are 0.1347, 0.0371, 0.1926, and 0.9619. The value on the proposed Multi-Perspective Fusion Bi-LSTM with Universal Scale CNN Model is 0.0908, 0.016, 0.1264, and 0.9836. When compared to other models the proposed model obtains the least error value. The graphical representation on table is depicted in figure 6.

5. CONCLUSION AND FUTURE RECOMMENDATION

Sales forecasting is one on the significant tasks in the dairy industry which assists to avoid the revenue loss and learn about the high sale locations. Many existing studies have employed various algorithms to forecast dairy sales; however, these models have encountered certain limitations. To overcome the problems, the proposed research has employed Multi-Perspective Fusion Bi-LSTM with Universal Scale CNN model to forecast the dairy sales. The proposed research has been assessed using the dairy goods sales dataset. The proposed Multi-Perspective Fusion Bi-LSTM with Universal Scale CNN model has compared with four regression models namely RF Regressor, linear regression, SVR, and DT Regressor. The performance on the proposed Multi-Perspective Fusion Bi-LSTM with Universal Scale CNN model has assessed using performance metrics on RMSE, MAE, and MSE. The proposed model has obtained 0.0908, 0.016, 0.1264, and 0.9836. In contrast to existing models, the proposed Multi-Perspective Fusion Bi-LSTM with Universal Scale CNN model has attained less error. The completed results on the proposed Multi-Perspective Fusion Bi-LSTM with Universal Scale CNN is very motivating, and promising. In future, the proposed research will analyze the products on MEVGAL (Greek dairy production company) to find out the external factors which impact the dairy market, technological, economic, and environmental factors. Additionally, it aids in understanding the current situation on the dairy industry and analyzing the prospective chances for growth and development.

AUTHOR CONTRIBUTIONS

Naveen D. Chandavarkar and Soumya S made substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data; or the creation of new software used in the work; Naveen D. Chandavarkar and Soumya S drafted the work or revised it critically for important intellectual content; Naveen D. Chandavarkar and Soumya S agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. Naveen D. Chandavarkar and Soumya S approved the version to be published.

References

[1] T. Musora, Z. Chazuka, A. Jaison, J. Mapurisa and J. Kamusha, "Demand Forecasting of a Perishable Dairy Drink: An ARIMA Approach," *International Journal of Data Mining & Knowledge Management Process*, vol. 13, no.1/2, pp.35-49, 2023.

[2] S. N. Raut, "Comparative analysis of Time Series Models to predict the seasonal prices of Ireland Dairy Products," M.S. thesis, Dublin, National College of Ireland, 2022.

[3] D. T. Tran, J.-H. Huh and J.-H. Kim, "Building a Lucy hybrid model for grocery sales forecasting based on time series," *The Journal of Supercomputing*, vol. 79, no. 4, pp. 4048-4083, 2023.

[4] J. Laktena, "Balancing Demand and Supply Planning in the Food Supply Chain: Insights from a Dairy Industry Case Study," M.S. thesis, Stockholm, KTH Royal Institute of Technology, 2023.

[5] C. Vithitsoontorn and P. Chongstitvatana, "Demand Forecasting in Production Planning for Dairy Products Using Machine Learning and Statistical Method," *2022 International Electrical Engineering Congress (iEECON)*, Khon Kaen, Thailand, pp. 1-4, 2022.

[6] G. K. Sinha and S. Mishra, "Sustainable Supply Chain Management Practices in the Dairy Industry: A Comparative Study of Leading Dairy Firms and Future Research Directives," Asian Journal of Dairy and Food Research, vol. 42, no. 4, pp. 435-446, 2023.

[7] T. S. Fagle, "Key Performance Indicators to Increase Logistics Performance in The Dairy Industry: A Case Study in The Dairy Industry," M.S. thesis, Stockholm, KTH Royal Institute of Technology, 2023.

[8] E. Delaney, D. Greene, L. Shalloo, M. Lynch and M. T. Keane, "Forecasting for sustainable dairy produce: enhanced long-term, milk-supply forecasting using k-NN for data augmentation, with prefactual explanations for XAI," in *International Conference on Case-Based Reasoning*, Springer, pp. 365-379, 2022.

[9] S. Wahyudi and M. Asrol, "Designing A Supplier Evaluation Model in The Cheese Industry Using Hybrid Method," *Academic Journal of Manufacturing Engineering*, vol. 20, no. 2, pp. 27-35, 2022.

[10] X. Ji, J. Wang and Z. Yan, "A stock price prediction method based on deep learning technology," *International Journal of Crowd Science*, vol. 5, no. 1, pp. 55-72, 2021.

[11] R. Huerta-Soto *et al.*, "Predictable inventory management within dairy supply chain operations," *International Journal of Retail & Distribution Management*, vol. 53, no. 3, pp. 1-17, 2023.

[12] G. Suseendran and B. Duraisamy, "Predication of dairy milk production using machine learning techniques," *Intelligent Computing and Innovation on Data Science*, pp. 579-588, 2021.

[13] V. Kumar, R. Sharma and P. Singhal, "Demand forecasting of dairy products for amul warehouses using neural network," *International*

Journal of Science and Research (IJSR), pp. 9-15, 2019.

[14] T. T. M. Ho, L. V. Tran, H. M. Tran and S. V.T. Dao, "Machine Learning in Demand Forecasting," *International Research Journal of Advanced Engineering and Science*, vol. 7, no. 3, pp. 225-233, 2022.

[15] T. Falatouri, F. Darbanian, P. Brandtner and C. Udokwu, "Predictive analytics for demand forecasting–a comparison of SARIMA and LSTM in retail SCM," in *3rd International Conference on Industry 4.0 and Smart Manufacturing*, vol. 200, pp. 993-1003, 2022.

[16] A. Atalan, "Forecasting drinking milk price based on economic, social, and environmental factors using machine learning algorithms," *Agribusiness*, vol. 39, no. 1, pp. 214-241, 2023.

[17] Z. Li, A. Zuo and C. Li, "Predicting Raw Milk Price Based on Depth Time Series Features for Consumer Behavior Analysis," *Sustainability*, vol. 15, no. 8, p. 6647, 2023.

[18] J. Malczewski and W. Czubak, "Hybrid Convolutional Neural Networks Based Framework for Skimmed Milk Powder Price Forecasting," *Sustainability*, vol. 13, no. 7, p. 3699, 2021.

[19] S. Yadav *et al.*, "Modeling and forecasting of producer price index (PPI) of cheese manufacturing industries," *Journal of Agriculture, Biology and Applied Statistics*, vol. 1, no. 1, pp. 39-49, 2022.

[20] Y. E. Gür, "Innovation in the dairy industry: forecasting cow cheese production with machine learning and deep learning models," *International Journal of Agriculture Environment and Food Sciences*, vol. 8, no. 2, pp. 327-346, 2024.

[21] N. Shyian, V. Moskalenko, O. Shabinskyi and V. Pechko, "Milk price modeling and forecasting," *Agricultural and Resource Economics: International Scientific E-Journal*, vol. 7, no. 1, pp. 81-95, 2021.

[22] S. Kresova and S. Hess, "Determinants of Regional Raw Milk Prices in Russia," 2021.

[23] A. Bełdycka-Bórawska *et al.*, "Price changes of dairy products in the European Union," *Agricultural Economics*, vol. 67, no. 9, pp. 373-381, 2021.

[24] B. G. Hansen, Y. Li, R. Sun and I. Schei, "Forecasting milk delivery to dairy–How modern statistical and machine learning methods can contribute," *Expert Systems with Applications*, vol. 248, p. 123475, 2024.

[25] A. N. Rezitis and G. Kastner, "On the joint volatility dynamics in international dairy commodity markets," *Australian Journal of Agricultural and Resource Economics*, vol. 65, pp. 704-728, 2021.

[26] A. N. Rezitis, O. A. Tremma, "The linkage between international dairy commodity prices and volatility: a panel-GARCH analysis," *Journal of Agribusiness in Developing and Emerging Economies*, vol. 13, no. 5, pp. 685-705, 2023.

[27] Q. Wang, R. Cheng and W. Xu, "Assessing volatility spillover effect between international milk powder and China's raw milk markets in the context of import growth," *Cogent Food & Agriculture*, vol. 9, no. 1, p. 2253715, 2023.

[28] S. Kresova and S. Hess, "Identifying the determinants of regional raw milk prices in Russia using machine learning," *Agriculture*, vol. 12, no. 7, p. 1006, 2022.

[29] C. G. Frasco *et al.*, "Towards an Effective Decision-making System based on Cow Profitability using Deep Learning," in *Proceedings of the 12th International Conference on Agents and Artificial Intelligence (ICAART 2020)*, vol. 2, pp. 949-958, 2020.

[30] A. Liseune, M. Salamone, D. Van den Poel, B. Van Ranst and M. Hostens, "Leveraging latent representations for milk yield prediction and interpolation using deep learning," *Computers and Electronics in Agriculture*, vol. 175, p. 105600, 2020.

[31] S. Badruddoza, A. C. Carlson and J. J. McCluskey, "Long-term dynamics of US organic milk, eggs, and yogurt premiums," *Agribusiness*, vol. 38, no. 1, pp. 45-72, 2022.

[32] M. Alwadi, A. Alwadi, G. Chetty and J. Alnaimi, "Smart dairy farming for predicting milk production yield based on deep machine learning," *International Journal of Information Technology* , vol. 16, pp. 4181-4190, 2024.

[33] S. V. Shankar *et al.*, "Modeling and forecasting of milk production in the western zone of Tamil Nadu," *Asian Journal of Dairy and Food Research*, vol. 42, no. 3, pp. 427-432, 2023.

[34] R. U. Mhapsekar, L. Abraham, S. Davy and I. Dey, "Application Adaptive Light-Weight Deep Learning (AppAdapt-LWDL) Framework for Enabling Edge Intelligence in Dairy Processing," in *IEEE Transactions on Mobile Computing*, vol. 24, no. 2, pp. 1105-1119, Feb. 2025.

Naveen D. Chandavarkar a Research Scholar at the institute of Computer Science, Srinivas University, Mangalore and serves as an Assistant Professor at Government First Grade College, Lingasugur, Raichur District, India. With 15 years of teaching experience, he specializes in software engineering and data mining. He has made notable contributions to academia through over 14 publications in refereed journals, showcasing his expertise and dedication to advancing his research fields. His work reflects a strong commitment to both education and scholarly innovation, enriching the academic community with his insights and practical knowledge in computer science.

Soumya S based in Mangaluru, India, is an accomplished academician with over 16 years of teaching and research experience in computer science. She holds an M.C.A and M.Phil., and a Ph. D in energy – efficient protocols for adhoc networks. Currently an Assistant Professor at Srinivas University. She specializes in Ad-Hoc networks, Wireless Sensor Networks and Artificial Intelligence. Dr Soumya has published 24 peer reviewed articles, authored two books, mentored 10 doctoral Scholars and registered 3 patents. Her Ph.D thesis earned the Best Thesis Award – 2022 and she has contributed extensively through conferences, journals articles and academic, governance, exemplifying her dedication to education and innovation.