Fusion Modeling for Worker-Centric Defect Analysis: Integrating Traditional and Time-Series Approaches in Apparel Manufacturing

Minuli Samaraweera
Department of Computer Science
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
minusamaraweera@gmail.com

Kavindu Kasthurirathne
Department of Computer Science
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
kavindu.kasthu@gmail.com

Minindu Thilakarathne
Department of Computer Science
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
mininduthilakarathne123@gmail.com

Chehan Weeransinghe
Department of Computer Science
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
chehanwe1998@gmail.com

Samadhi Rathnayake
Department of Computer Science
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
samadhi.r@sliit.lk

Supipi Karunathilake
Department of Computer Science
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
supipi.k@sliit.lk

***Abstract*—SeamSense was developed to significantly improve defect rate prediction in the apparel manufacturing sector by integrating advanced time-series analysis, machine learning, and worker-centric data. This innovative system leverages both historical and real-time data, enabling precise and reliable defect forecasts. By incorporating worker demographic information, SeamSense not only enhances predictive accuracy but also provides actionable insights into the root causes of defects, allowing for targeted interventions. The system's ensemble modeling approach combines multiple model outputs, enhancing robustness and adaptability to changing manufacturing conditions. SeamSense has proven effective in dynamic production environments, demonstrating its potential applicability across various industries. Future developments will focus on increasing its versatility, aiming to extend its use to other sectors, thus contributing to broader advancements in quality management and predictive analytics.**

Keywords— Defect prediction, apparel manufacturing, time-series analysis, machine learning, worker-centric data, fusion model, ensemble modeling, quality control.

I. INTRODUCTION

Quality control and reduction of defects are more critical today than ever before due to the dynamics of today's manufacturing world. Obstacles surrounding the management of defects are apparent in the global apparel sector due to multifaceted processes and reliance on human workforces. Fluctuations in worker performance, material quality, and environmental conditions, thereby makes forecasting and controlling of defects passive. However, in this complex environment, the performance of most of the conventional defect prediction models, which are mainly based on statistical models and past data, are generally unsatisfactory.

There is also potential for bettering the existing approaches in defect prediction due to the new technologies offered by the machine learning and data analytics. Time series forecasting has been the most common technique in predictive modeling since it uses previous data to establish patterns. But this method by itself is not enough to capture the complex interdependencies many factors that may influence the defect rates. especially when human aspects such as the attributes of the workers are relevant. This is why it is necessary to develop models that would also consider context and the temporal nature of data to overcome the limitations.

Introducing the "Fusion Modeling for Worker-Centric Defect Analysis" systematic approach is modern regarded machine learning and survivor statistical models. This system offers a comprehensive structure of a defect prediction system by employing an analysis of time series, comprehensive workers' profile metrics, and constant production monitoring. From the point that a variety of sources of information can be included into a single model, a more extensive understanding of the patterns of faults occurs, which enables manufacturers to predict and address potential issues.

This system can yield improvements in total production quality, from an increase in the precision of predictions about defects to offering practical insights. This approach is better than traditional approaches because it addresses the role of human factors in the rate of defects and offers an in-depth understanding of the manufacturing process. The more far-reaching implications of the research suggest that quality management and defect prediction are significant for a range of businesses. We believe that this research will provide a firm framework for defect analysis by coupling real-time data processing with advanced predictive algorithms. It lays the groundwork for further advances in the industry by contributing to the ongoing discussion on how best to use machine learning and data analytics for better industrial outcomes.

II. RELATED WORK

Thorough research has been conducted on the ability to reliably predict manufacturing problems, especially in businesses where maintaining good product quality is essential. Factors including worker variability, material inconsistencies, and environmental impacts increase the issue of defect management in the garment business, where production processes are complex and strongly dependent on human labor. Although they have provided valuable insights, traditional approaches to defect prediction—which are frequently based on statistical analysis and time-series forecasting—are unable to fully address the complex structure of modern manufacturing processes.

In manufacturing, time-series analysis has long been an essential part of predictive modeling. Since they are excellent at predicting trends based on past data, methods like ARIMA (AutoRegressive Integrated Moving Average) are frequently employed to represent seasonal trends in defect data. Research like the ones by [1] [2] have shown how well these models work with linear and stationary data sets. But these conventional models have trouble with non-linearities and the dynamic nature of manufacturing processes, especially in situations where human elements are present [2].

Defect prediction capabilities have advanced significantly with the introduction of machine learning. More and more machine learning models, such as deep learning algorithms, decision trees, and support vector machines (SVMs) are being used in manufacturing settings to anticipate failures more accurately. The textile industry's success with random forest algorithms was highlighted in [3], demonstrating how machine learning may detect complicated patterns that conventional statistical techniques can overlook. Considering these developments, there is still more work to be done in order to fully realize the promise of these models by including human-centric data [4].

In light of the significance of human variables in the occurrence of defects, worker demographics have started to be included in predictive models in recent studies. The important influence that worker personality traits and demographics have on system performance is highlighted by research by [5]. By integrating these elements, predictive models can achieve a more nuanced understanding of defect causality, leading to more targeted interventions and improved quality control [6].

Even with these developments, there is still a significant clean in the literature: there are no comprehensive models that successfully integrate time-series analysis, machine learning, and worker-centric data. Current models frequently consider every input to be equally important, which may cause them to ignore the crucial impact of worker experience, skill level, and other demographic characteristics. This research gap is especially observable in the apparel industry because of the wide range of labor and the production process [7].

Fusion modelling is now recognized as an effective way for filling these gaps. To improve overall accuracy, this approach combines several prediction models, such as machine learning algorithms with conventional time-series models. Increasingly, strategies like stacking, boosting, and bagging are being utilized to combine the advantages of many models to increase predictive performance. presented a multivariate prediction model that greatly increases accuracy in difficult data scenarios by combining ensemble learning and three-way clustering. These methods are especially pertinent in the manufacturing sector, where complex and non-linear interactions can occur between a variety of elements, including materials, machines, and human operators [8].

Fusion modeling has gained popularity as an approach for defect prediction. It involves combining multiple predictive models to increase overall accuracy. It has been demonstrated that ensemble techniques, including as stacking, boosting, and bagging, improve prediction performance by utilizing the advantages of various algorithms. For instance, [9] presented a multivariate prediction model that considerably increases prediction accuracy in complicated data settings by combining ensemble learning with three-way clustering. These methods are especially useful in manufacturing, where complex and non-linear interactions can occur between a variety of components, including materials, machines, and human operators. Fusion modeling can offer a more comprehensive understanding of the reasons causing failures by combining the outputs from several models, which can result in more successful interventions and process changes.

The study focuses on minimizing defects in the sewing department of apparel manufacturing, a critical aspect in our research on defect prediction. It underscores the challenges faced due to human error and the need for continuous quality improvement. [11]The emphasis on statistical analysis and root cause analysis aligns with our objective of developing predictive models to identify and prevent defects. The discussion on the consequences of defects, such as wasted production costs, highlights the significance of defect prediction and prevention. Keywords like "Quality control," "Data analysis," and "Defects" resonate with our research goals, emphasizing the relevance of this study to our component.

The proposed "Fusion Modeling for Worker-Centric Defect Analysis" system aims at building upon these developments by fusing worker demographic data with time-series analysis. By using a fusion method, defect predictions become more robust and offer a more thorough insight of the elements influencing manufacturing problems. In order to close the gaps and advance the field of analytics for prediction in quality management, this project will make use of both conventional statistical approaches and modern facilities machine learning techniques.

III. METHODOLOGY

This research methodology lays special emphasis on integrating advanced machine-learning techniques with conventional time-series analysis to enhance defect prediction in the garment manufacturing industry. The study seeks to confirm the reliability and validity of the findings by systematically collecting, analyzing, and decoding information on worker demographics and occurrences of defects.

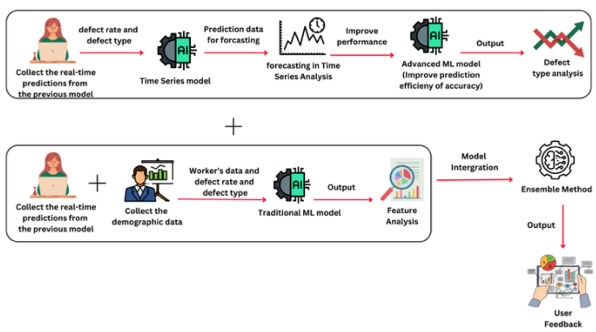

Fig. 1. Fusion Model for Defect Analysis Diagram

In SeamSense, the fusion model is presented as a new concept of defect analysis that utilizes both machine learning and time series analysis. This combined approach combines the advantages of both methodologies in the analysis of defect patterns.

Above **Figure 1** presents the fusion model workflow. It starts with capturing defect rate and worker demographic data then it involves both time series and other conventional machine learning techniques. The results obtained from the aforementioned models are, therefore, recombined using an ensemble technique that improves the defect type classification. The effectiveness of the model is constantly enhanced in light of user feedback thus guaranteeing efficiency in a volatile manufacturing setting.

1) Data Collection and Preprocessing: A large-scale apparel was the study's main source of data. The collection contains historical defect rates, real-time production indicators, and detailed worker demographic data, including work hours, experience, and skill level. Over the course of six months, this data was collected in order to guarantee an accurate understanding of the factors impacting the occurrence of defects. In order to obtain accurate and current data, the gathering method used both manual logging systems and automated sensors.

2) Data Analysis: Using the ARIMA model, time-series analysis was applied to the gathered data in order to find patterns in the temporal incidence of defects. The efficacy of this conventional statistical technique in predicting trends from past data led to its selection. In parallel, worker demographics and defect rates were analyzed using machine learning methods like Random Forest and SVM. To guarantee their accuracy and durability, these models were cross-validated after being trained on a subset of the data.

3) Traditional Machine Learning: Traditional machine learning models, including Random Forest, Support Vector Machines (SVM), and Lasso Regression, were employed to analyze the relationship between worker demographics and defect rates. These models were trained on a subset of the data and validated using cross-validation techniques to ensure accuracy and robustness. The traditional models focus on static features such as worker demographics, machine settings, and environmental conditions, providing valuable context to the prediction of defect rates. [9].

4) Time-Series Analysis: The periodic changes in defect rates are the main focus of time-series analysis. The approach can detect trends, seasonality, and recurrent trends in historical data that might not be seen with conventional analysis alone. When it comes to capturing the dynamic character of the production process and understanding how defect rates change over time, time-series analysis is very useful.

5) Fusion Model Implementation: The fusion model, which combines the outputs from the time-series and conventional machine learning models, represents the innovation in this study. By combining the best features of the two methodologies, this hybrid model develops a thorough framework that takes into account both static and dynamic factors affecting failure rates. Quality control and process optimization benefit greatly from the integration, which also improves prediction accuracy and offers a comprehensive picture of problem occurrence.

6) Model Validation and Testing: To assess the predictive models' reliability and precision, they conducted extensive testing using a different validation dataset. To assess the efficacy of the fusion modeling technique, performance metrics including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) were computed. To confirm the models' suitability for dynamic manufacturing environments, real-time testing was also conducted in the production setting.

7) Model Development: SeamSense utilizes a hybrid approach, developing separate models for time-series forecasting and machine learning analysis. Based on past trends, time-series models like ARIMA or Prophet are used to forecast future trends in defects. Meanwhile, classic machine learning models such as Random Forest or XGBoost—are being built to examine how worker demographics affect defect rates. To guarantee the robustness and dependability of these models, cross-validation techniques are used during training and validation.

8) User Feedback and Continuous Improvement: Outline the feedback loop from production managers and quality control teams, which helps refine and improve the model over time

9) Addressing Potential Limitations: The study has identified worker performance variability as one potential limitations that may bring noise into the dataset. Strong data preparatory treatment methods, such as outlier detection and normalization, were used in the study to guarantee the data's integrity in order to lessen this. Additionally, the fusion modeling approach's iterative nature enables ongoing enhancement of models as new data becomes available, thereby increasing the models' accuracy and adaptability.

These two methods are used in the fusion model to provide a comprehensive understanding of defect frequency. This integrated technique considers both static and dynamic elements, improving defect detection accuracy and effectiveness. As a result, there is a stronger and more reliable defect analysis model that may offer insightful information about process optimization and quality control in a worker-centric manufacturing setting.

IV. Results and Discussion

The study was on the application of different traditional machine learning techniques to predict the defect rate in manufacturing for four specific defect types: Run Off, Open Seam, SPI Errors, and High Low. The models used are Random Forest, Gradient Boosting, Support Vector Regressor, Linear Regression, Ridge Regression, and Lasso Regression. The best among the defect types was Lasso Regression overall, as it selected features in the most efficient way without overfitting. The model greatly improved in predicting Open Seam and Run Off defects. Particularly robust and effective was the Lasso Regression model upon parameter tuning using RandomizedSearchCV and with good performance metrics. Lasso Regression gives an increased R^2 score, showing the model explains a good variation of defect rates, especially in cases of Open Seam and Run Off. Even though the prediction is strong by the model for most defect types, it still requires a bit more fine-tuning to predict the SPI Error and High Low defects accurately.

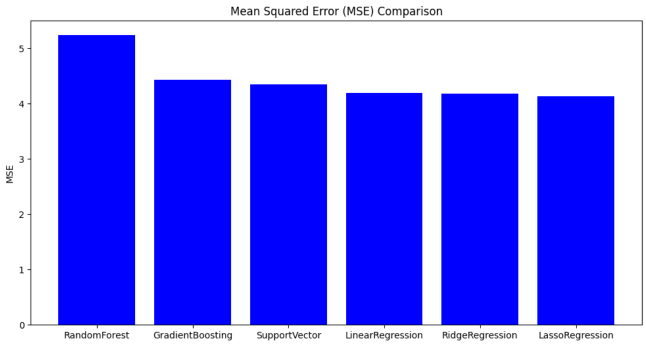

Fig. 2. Model Comparison

Lasso Regression performs better than other typical machine learning models, obtaining the lowest Mean Squared Error (MSE) among the models that were tested. This suggests that Lasso Regression, especially for Open Seam and Run Off flaws, offers the most precise estimates for defect rates. Random Forest, on the other hand, showed the greatest MSE, indicating a lesser level of prediction accuracy. Overall, the findings point to Lasso Regression as the best model for this study, providing a dependable approach to manufacturing defect prediction and quality management.

The Lasso Regression model had an MSE of 4.014, and its R^2 performance was consistent over various defect types. From this consistency, it is evident that the model was quite effective at explaining how the demographics of workers influenced defect rates. The correctness of the model further stayed in the affirmative by the MAE, which reiterated that such a model, like Lasso Regression, was appropriate for this type of scenario, especially when there is a precise need to identify defects, particularly in cases with low variance in defective items.

TABLE I. TIME SERIES MODEL PERFORMANCE

Defect Type	**Measures**			
	RMSE	***F1 Score***	***Accuracy***	***Precision***
Run Off	4.34	0.88	0.8	0.8
Open Seam	1.56	0.88	0.8	0.8
SPI Error	2.24	0.75	0.6	0.6
High Low	1.90	0.88	0.8	0.8

Finally, we present the performance of ARIMA models in predicting the occurrence of defects over a period of time. In general, the ARIMA models have shown good scores, especially in precision for SPI Errors and High Low defects, where detection is important, in terms of F1 scores and Accuracy. However, with Open Seam and Run Off, higher RMSE values were revealed, making the forecast for those defects much more challenging, perhaps because the temporal patterns are more complex or are affected by other external factors that were not considered in the model.

A stacking ensemble has been built so that the strengths of both the Lasso Regression and ARIMA models can be leveraged. The effectiveness of this proposed fusion model is demonstrated to outperform other individual models for the defects Open Seam and Run Off in terms of both accuracy and prediction. Pooling results together from static and dynamic factors caught by the Lasso Regression and ARIMA models makes this model more comprehensive and adaptive toward the predictive framework. This is evidenced by the higher R^2 values and the smaller MAE values of the fusion model compared to those of the individual models. Therefore, this points out the strength of the fusion model in quality control and defect prevention within dynamic manufacturing environments. The result is such that Lasso Regression turns out to be more precise for applications at higher resolutions demanding exact counts of defects, and in the binary classification scenario, ARIMA models outstrip Lasso Regression.

Efforts were made to compare the prediction of the defect rates in various manufacturing scenarios using the presented models above and hence to answer which model is considered best for the prediction of defects like Open Seam and Run Off. The performance of the model was less robust with defects such as SPI Errors and High Low, whereby more fine-tuning is needed. The fusion model is capable of capturing both static and dynamic factors correctly and using Lasso Regression and ARIMA models through the stacking ensemble approach for very accurate predictions, which could adapt more to the manufacturing environment. The fact that the model accuracy keeps improving for all defect categories testifies to the applicability of the fusion model in practical, evolving manufacturing setups. Also, the high R^2 values and low MAE further prove that the fusion model is effective in quality control and defect prevention.

In subsequent research, we should go a step ahead in exploring more advanced ensemble methodologies—for example, boosting or bagging—since perhaps they could assist in pushing the predictive accuracy still higher. Similarly, hybrid models combining machine learning with deep learning techniques, such as CNNs with ARIMA, may also be developed for removing the limitations of the above models. A need also exists to explore generic adaptations of the fusion model for applications across different industrial sectors or manufacturing scenarios, and specific modifications are suggested by the review. Industry-specific case studies would further validate and provide deeper insight into the general applicability of the model.

A critical aspect of applying these models in real life is addressing potential biases in processing worker demographic data. The research highlights strategies to minimize biases, ensuring fair, transparent, and interpretable predictions. Techniques like SHAP values are proposed to provide clear explanations for the model's predictions, fostering trust in the final decisions.

A. Real-World Deployment Considerations

A few key factors must be considered in the deployment of the fusion model within real manufacturing settings: The model should be very extensible and handle real-time data coming from multiple sources. To maintain optimal performance, it should support distributed computing or cloud-based infrastructure, which can scale as needed. This enables integration with Manufacturing Execution Systems (MES) or Enterprise Resource Planning (ERP) systems so that predictions of defects are delivered in an easily accessible and actionable manner for both production managers and quality control teams. An intuitive interface guides these teams in understanding results, searching for patterns, and making timely decisions.

Moreover, the developed model should be made to be capable of learning iteratively as new data comes in, along with periodic updates, reflecting the most recent manufacturing conditions. This needs self-organized pipelines for continuous evaluation and optimization for the efficiency of the model. The data protection and privacy provisions should include encryptions and all other security protocols that have been outlined in industry standards. Furthermore, it is important to take into consideration the cost-effectiveness; the benefits of deploying the model should not cause overbearing costs in terms of achieving a positive return on investment.

It is essentially the key that workers buy in to allow its implementation. The core thing for its effective operationalization is training and engagement of the workforce. Real-time monitoring and alert systems could allow a proactive defect management system, thus reducing production downtime. Finally, ethical considerations should be first and foremost to ensure that the model supports human decision-making, is transparent in operation, and remains free from bias that could skew results.

V. Conclusion

One milestone in developing SeamSense was to analyze manufacturing defect rates, especially in industries where human factors and minor fluctuations in productivity are critical. By integrating advanced time-series analysis, machine-learning techniques, and a rich tapestry of worker demographic information, SeamSense enables identification of problems and correction of them, efficiently and effectively. This holistic view not only brings greater precision in predicting defects but also presents manufacturers with practical tools for robust quality control measures.

One of the amazing features of SeamSense is its adaptability to new and changing needs of the production lines. This is maintained through iterative processes of model refinement that are continuously informed by real-time feedback. In modern set-ups, a system's capability to integrate real-time data is updated to result in performance dynamically with changes in the production environment.

From the above applications in the apparel industry, SeamSense can later become an effective software for total quality management for other industries as well. Besides the apparel industry, SeamSense could improve its predictive analytics and process enhancement in technology and automotive manufacturing by improving its models to capture a wider array of production data. This research not only significantly pushes the boundaries of current defect prediction practice but also lays a foundation for future studies aimed at improving production efficiency and quality in diverse industries.

References

[1] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," January 2003. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0925231201007020. [Accessed August 2024].

[2] B. C. and S. K. D. , "Modeling and Analysis of Defect Data—A Time Series Approach," September 2022. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-19-2188-9_85. [Accessed August 2024].

[3] "Minimization of Defects in the Sewing Department,," 2019. [Online]. Available: http://14.139.111.20:7888/jspui/handle/1/886. [Accessed February 2024].

[4] B. Pavlyshenko, "Using Stacking Approaches for Machine Learning Models," August 2018. [Online]. Available: https://vpn.sliit.lk/proxy/031b55e2/https/ieeexplore.ieee.org/document/8478522.

[5] D. C. J. and . L. W. S. , "Using worker personality and demographic information to improve system performance prediction," August 2004. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0272696304000439. [Accessed August 2024].

[6] R. R. M. D. N. R. S. P. M. and A. B. E. , "An ML-Based Approach for Optimizing the Productivity and Efficiency of the Apparel Industry by Focusing on Trainee Employees," October 2023. [Online]. Available: https://www.proquest.com/docview/2896757871?fromopenview=true&pq-&sourcetype=Scholarly%20Journals.

[7] M. H. D. M. R. and L. d. S. C. , "Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series," January 2020. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1568494619306180. [Accessed May 2024].

[8] X. W. Wu, J. Zhan and W. Ding, "TWC-EL: A multivariate prediction model by the fusion of three-way clustering and ensemble learning," December 2023. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1566253523002828. [Accessed July 2024].

[9] "Sri Lanka Export Development Board (EDB)," Linea Aqua (Pvt) Ltd, [Online]. Available: https://www.srilankabusiness.com/exporters-directory/company-profiles/linea-aqua-pvt-ltd/.

[10] G. B. S. B. Taieb and Y.-A. L. B. , "Machine Learning Strategies for Time Series Forecasting," [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-36318-4_3. [Accessed June 2024].

[11] M. Seçkin, A. Ç. Seçkin and A. Coşkun, "Production fault simulation and forecasting from time series data with machine learning in glove textile industry," October 2019. [Online]. Available: https://journals.sagepub.com/doi/full/10.1177/1558925019883462. [Accessed July 2024].

[12] "Wikipedia," MAS Holdings, 31 January 2024. [Online]. Available: https://en.wikipedia.org/wiki/MAS_Holdings.

[13] "MAS Holding," [Online]. Available: https://masholdings.com.

[14] A. M. -Esteban , F. M.-Á. A. T. J. J. and C. R.-E. c. , "Pattern recognition to forecast seismic time series," December 2010. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0957417410004616. [Accessed July 2024].

[15] A. P. C. and A. K. G. , "Ensemble Machine Learning Approach for Detecting and Predicting Diabetes Mellitus Using Bagging and Stacking," December 2023. [Online]. Available: https://vpn.sliit.lk/proxy/031b55e2/https/ieeexplore.ieee.org/search/searchresult.jsp?newsearch=true&queryText=Using%20Stacking%20Approaches%20for%20Machine%20Learning%20Models. [Accessed Juy 2024].

[16] R. S. P. J. B. P. L. and G. G. , "Multistage Quality Control Using Machine Learning in the Automotive Industry," June 2019. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8737933. [Accessed July 2024].

[17] W. L. X. Y. Q. Z. G. C. X. H. and S. H. , "Time Series Forecasting Fusion Network Model Based on Prophet and Improved LSTM," August 2022. [Online]. Available: https://www.techscience.com/cmc/v74n2/50249. [Accessed July 2024].