**INFO-B 691: Project/Thesis Pre-Assessment Form**

**Student Name:** \_\_\_\_\_\_\_\_\_\_Puja Darshana Mishra\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Student ID**: \_\_\_\_\_\_2001290409\_\_\_\_\_\_\_\_\_\_\_

**Project**

**Faculty advisor: \_\_\_\_\_Saptarshi Purkayastha\_\_\_\_\_\_\_\_\_\_\_**

**Semester to be enrolled in: Spring 2025**

**Personalized Prognosis for Chronic Wound Healing: Predicting Healing Time Using a Multivariate Approach**

**1. What are your personal and professional goals for this project/thesis?**

Personal goals: My personal goal is to work on an independent capstone project and contribute to the development of personalized healthcare decision support tools using machine learning approaches.

Professional goals: My professional goal is to be able to apply an advanced machine learning techniques to derive meaningful insights from multivariate datasets, combining demographic, clinical data and treatment histories. This would equip me with experience in creating a system that can be practically used in clinical settings.

**2. Give a detailed description of your project/thesis.**

* **Introduction:**

This project focuses on creating a personalized prognosis system for predicting the healing time of chronic wounds based on individual patient factors. Chronic wounds are the wounds that fail to heal in the normal path in a timely manner, the reparative process is delayed due to various factors. These are the wounds that fail to produce anatomic and functional integrity over a period of 3 months (Boukovalas et al., 2023). The model will incorporate multivariate data such as patient demographics (age, gender, comorbidities), Wound characteristics (depth, stage, infection, necrosis, wound size, granulation, fibrin), and Treatment data (medication, compliance). Using this data, a machine learning model will be developed to predict healing time in days and provide a confidence interval (uncertainty estimates) around that prediction. This effort aligns with current advancements in personalized medicine, emphasizing the use of machine learning and predictive analytics to improve patient-specific care (Veličković et al., 2023).

* **Knowledge Gap in the area:**
* **Lack of Personalization:**  
  Current clinical guidelines often apply population-level recommendations, which fail to address individual variability in wound healing outcomes (Bender et al., 2021). The factors that affect wound healing are very subjective, which calls for a need to develop personalized models for the patients.
* **Data Complexity and Heterogeneity:**  
  Chronic wounds are influenced by multiple factors, including patient demographics, comorbidities, wound characteristics, and treatment regimens. Managing this heterogeneity requires advanced data-driven approaches.
* **Uncertainty in Predictions:**  
  Existing models rarely account for uncertainty, which is critical in clinical decision-making. For instance, clinicians often lack tools to gauge confidence intervals for predicted recovery timelines (Krauth et al., 2016). This project aims to employ multiple models to understand the accuracy of prediction.
* **Limited Integration with Standardized Clinical Documents:**

Many existing wound care tools lack the capability to directly process structured clinical documents such as Continuity of Care Documents (CCDs) and Clinical Document Architecture (CDA) files, which are widely used for storing and sharing healthcare information (Perugu et al., 2023). This project addresses this gap by designing a system that can import and analyze CCD/CDA files, extracting relevant patient and wound-related data automatically.

* **Aim**

This project aims to develop a machine learning-based system to predict healing time in days based on various patient factors that are commonly recorded for patients undergoing treatment for chronic wounds. When integrated in Clinical decision-making tools, the project will equip clinicians with actionable insights into patient recovery timelines, improving decision-making. It provides patients with transparent, understandable recovery forecasts, fostering trust and compliance with treatment plans.

* **Objective:**

The objective of this project is to develop a system that utilizes machine learning to integrate various types of data like permanent (e.g., demographics), semi-permanent (e.g. comorbidities) and dynamic data (e.g. wound progression, moisture content in the wound) related to wound healing and gives accurate prediction of healing times based on this data.

To develop this system, several machine learning models will be implemented and compared to assess their performance on the dataset. A Transformer-based model, with a deep learning architecture, will be utilized for its ability to process data parallelly, this significantly speeds up the learning process due to reduced training and prediction time. Transformers can capture long-term dependencies in the data using their self-attention mechanism, this mechanism allows the model to weigh the importance of different points in the time series regardless of their distance in the input sequence, which is crucial for capturing seasonal trends and cyclic behaviour in time series data. They can incorporate contextual information through embeddings, thus enhancing their predictive accuracy, which could be particularly important for data on chronic wound and this project as it aims to personalize prediction for each patient. Transformers have been shown to have achieved They have achieved state-of-the-art results in several forecasting tasks, including the forecasting of influenza-like illnesses, demonstrating their capability to handle not only the temporal dynamics but also the complex interactions between different data dimensions (Wu et al., 2020, Zeng et al., 2023)​.

Gated recurrent units (GRU) have emerged to be an effective alternative to traditional models like LSTMs and will be utilized in this project. GRUs have a simpler structure compared to other recurrent neural networks like LSTMs, which makes them faster to train and handle large datasets. By using update and reset gates, GRUs can maintain long-term dependencies without gradients vanishing over time, which is crucial for learning from time series data that may have important dependencies at different time lags​ (Pavithra et al., 2019). Novel extreme adaptive GRU (eGRU), showcases their capability to adapt to different scenarios, including those where data points are outliers or represent rare, significant events this feature could be utilized to select features for personalization (Zhang et al., 2024). This is useful when past data is key to predicting future outcomes which is apt to this project.

An Autoregressive Integrated Moving Average (ARIMA) model, which is well-suited for time-series data, will model dependencies based on past observations, making it an appropriate choice for the temporal nature of wound care data and will serve as baseline for comparison.

* **Handling Uncertainty in Predictions:**

Uncertainty in predictions will be managed by addressing two key types of uncertainty:

* 1. Epistemic Uncertainty or Model Uncertainty:

Epistemic uncertainty represents the lack of knowledge within the model, typically stemming from limited or insufficient training data. It reflects uncertainty about the model’s parameters and can be reduced with additional or more diverse data. This uncertainty will be estimated using Monte Carlo Dropout, which approximates Bayesian inference by applying dropout during both training and inference. By executing multiple stochastic forward passes through the model and analyzing the variance in the predictions, the system will quantify epistemic uncertainty (Krauth et al., 2016).

* 1. Aleatoric Uncertainty or Data Uncertainty:

Aleatoric uncertainty arises from inherent noise or variability within the data itself, this will be captured by extending the model’s output to include not only the predicted healing time but also the variance associated with the prediction.

**3. What is the purpose of your project/thesis?**

The idea of this project is based on the growing importance of precision medicine, Weigelt et al. (2020) discuss how non-healing chronic wound have been a growing global health crisis, with mortality rates and management costs surpassing many cancers. They have emphasized that the need of the hour is to develop point-of care-diagnostic tools which are predictive, prescriptive, and personalized. Individualized risk management tools using machine learning have previously been utilized and have been found to be useful in chronic wound management (Velickovic et al., 2023). Peterson et al. (2017) have previously encouraged the idea of moving away from one-size-fits-all population models to personalized predictive models that adapt to each patient's unique clinical data over time to provide tailored predictions. They achieved this using a personalized Gaussian Processes (pGPs), which incorporate patient-specific data (e.g., cognitive test scores, imaging biomarkers, and demographics) to refine the population-level model for individual use for Alzheimer’s disease progression. The gap here is that these GPs can only work on data which is small and cannot predict long term variations or complex temporal relationships which are important with respect to the multivariate nature of factors affecting the progression of wound healing (Krauth et al., 2017). They also heavily rely on the quality and completeness of the historical data which further limits their use (Peterson et al., 2017). To overcome these shortcomings, a transformer model-based approach is adopted for this project as these models are good with large dataset and handling complex relationships and long-term variations. The GP models are good with calculating the uncertainty of prediction and to obtain this in a transformer model technique like Monte-Carlo dropout will be used, which will allow the model to generate multiple predictions for each input, capturing the variability in healing outcomes and providing confidence intervals. This will be an indication of how confident the mode is about these predictions. The variables in the study are taken from a previous study by Bender et al. (2021) where these variables proved to be most informative.

**4. Who is the target audience of your project/thesis? (Be specific)**

Primarily Clinicians, nurses, and wound care specialists who manage chronic wounds and make decisions about treatment plans and patient care**.** Secondarily, patients who are undergoing treatment and hospital administration in charge of maintaining resources and inventory necessary to provide timely care to these patients.

**5. State your expected outcomes or deliverables of this project/thesis.**

* A trained predictive model capable of estimating the healing time of chronic wounds based on patient-specific data.
* A comprehensive evaluation report detailing the model’s accuracy.
* Visualizations: Line charts and other visualizations comparing performance metrics across the different models, Visualization of the healing trajectory which shows the predicted trajectories of different models and a Feature importance plot showing important features determining healing time.
* An interactive application enabling users to upload CCD/C-CDA files, extract relevant clinical data, and map variables to the predictive model's features. The app will generate personalized healing time predictions with uncertainty intervals, enhancing clinical decision-making. Built using frameworks like Streamlit, it would ensure seamless integration into healthcare workflows.

**6. State expected timeline of your project/thesis. Be sure to include benchmark times that can be checked by faculty advisor.**

* Week 1-2: Literature review and data collection.
* Week 3-4: Data preprocessing and initial analysis, UI development
* Week 5-8: Model Training
* Week 8-10: Model testing and evaluation, fine-tuning for improvement.
* Week 11-12: UI implementation and app development
* Week 13: Report writing.

**7. Projected resources and or sources of information you will need to complete the project successfully.**

* Comprehensive dataset of anonymized wound healing and treatment outcomes.
* Software, programming and computational resources for model training and testing,

**8. Strategies of assessment and or usability studies. Detail how you plan on determining the success of this project/thesis.**

The evaluation of the model will include a comprehensive set of metrics to assess both statistical performance and clinical relevance**.**

1. **Mean Absolute Error (MAE):**

MAE measures the average magnitude of errors in predictions, providing a clear understanding of the typical prediction deviation in healing time. A lower MAE indicates that the model’s predictions are closer to the true values, which is essential for accurate treatment planning in wound care.

1. **Root Mean Squared Error (RMSE):**

RMSE is a standard metric for continuous prediction tasks, providing an important benchmark for model performance. It gives more weight to larger errors by squaring the differences between predicted and actual values before averaging. This metric is useful for understanding the overall error magnitude

1. **Time-Dependent Area Under the Curve (AUC):**

Time-dependent AUC will be calculated at various time horizons (e.g., 30, 60, 90 days) to assess the model’s ability to accurately predict healing trajectories over time. This metric evaluates the model’s performance in distinguishing between patients with different healing outcomes at various stages, providing a dynamic view of prediction accuracy across the healing process.

1. **Concordance Index (C-Index):**

The C-index will be used to evaluate the model's ability to rank patients based on predicted healing time, a higher C-index indicates that the model does a better job at ordering patients according to their healing progress. Predictions will also be evaluated against clinically relevant thresholds, such as whether the predicted healing time is within ±X days of the actual healing time.

**9. What new information do you think you will gain by doing this project/thesis?**

Practical experience in applying transformer-based models for real-world healthcare problems and learning how to deal with complex, multimodal patient data.

References:

Boukovalas, S., Aliano, K. A., Phillips, L. G., & Norbury, W. B. (2023). Wound healing. In C. Townsend, R. Beauchamp, B. Evers, & K. Mattox (Eds.), *Sabiston Textbook of Surgery* (21st ed., pp. 119–149). Elsevier.

Retrieved from: <https://www-clinicalkey-com.proxy.ulib.uits.iu.edu/#!/content/book/3-s2.0-B9780323640626000062?scrollTo=%23hl0001546>

Bender, C., Cichosz, S. L., Pape-Haugaard, L., Hartun Jensen, M., Bermark, S., Laursen, A. C., & Hejlesen, O. (2021). Assessment of simple bedside wound characteristics for a prediction model for diabetic foot ulcer outcomes. *Journal of diabetes science and technology*, *15*(5), 1161-1167.

Krauth, K., Bonilla, E. V., Cutajar, K., & Filippone, M. (2016). AutoGP: Exploring the capabilities and limitations of Gaussian process models. *arXiv preprint arXiv:1610.05392*.

Perugu, B., Wadhwa, V., Kim, J., Cai, J., Shin, A., & Gupta, A. (2023). Pragmatic approaches to interoperability–surmounting barriers to healthcare data and information across organizations and political boundaries. *Telehealth and Medicine Today*, *8*(4).

Peterson, K., Rudovic, O., Guerrero, R., & Picard, R. W. (2017). Personalized Gaussian processes for future prediction of Alzheimer's disease progression. *arXiv preprint arXiv:1712.00181*.

Veličković, V. M., Spelman, T., Clark, M., Probst, S., Armstrong, D. G., & Steyerberg, E. (2023). Individualized risk prediction for improved chronic wound management. *Advances in wound care*, *12*(7), 387-398.

Weigelt, M. A., Lev-Tov, H. A., Tomic-Canic, M., Lee, W. D., Williams, R., Strasfeld, D., ... & Herman, I. M. (2022). Advanced wound diagnostics: toward transforming wound care into precision medicine. *Advances in Wound Care*, *11*(6), 330-359.

Wu, N., Green, B., Ben, X., & O'Banion, S. (2020). Deep transformer models for time series forecasting: The influenza prevalence case. *arXiv preprint arXiv:2001.08317*.

Zeng, A., Chen, M., Zhang, L., & Xu, Q. (2023, June). Are transformers effective for time series forecasting?. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 37, No. 9, pp. 11121-11128).