

## MSc. in Computing Practicum Approval Form

---

Project Title:	Innovative Cirrhosis Analysis using predictive modelling.
Student ID:	23265480 & 23265467
Student name:	Sudharsan Swaminathan & Ezhil Nandhini Karthikeyan
Student email	<a href="mailto:sudharsan.swaminathan2@mail.dcu.ie">sudharsan.swaminathan2@mail.dcu.ie</a> , <a href="mailto:ezhil.karthikeyan2@mail.dcu.ie">ezhil.karthikeyan2@mail.dcu.ie</a>
Chosen major:	Data Analytics
Supervisor	Mark Roantree
Date of Submission	24-11-2023

**What is the topic of your proposed practicum?** (100 words)

Topic: **"Innovative Cirrhosis Analysis using Predictive Modelling"**.

This project aims to develop a predictive model for analysing cirrhosis, a type of liver disease, and to make clusters (K-medoids) based on clinical and demographic features. Leveraging unique values in categorical columns and transforming continuous variables into meaningful categories, we explore patterns indicative of cirrhosis disease. Through hypothesis testing and predictive analytics like Logistic Regression, and Random Forest, to seek enhanced accuracy of the disease, contributing to more informed clinical decision-making. This interdisciplinary approach with medical insights and ML techniques yields an understanding of this complex health condition.

**Please provide details of the papers you have read on this topic (details of 5 papers expected).**

1. Nusrat, Salman, Muhammad S Khan, Javid Fazili, and Mohammad F Madhoun. "Cirrhosis and Its Complications: Evidence Based Treatment." *World Journal of Gastroenterology*: *WJG* 20, no. 18 (May 14, 2014): 5442–60. <https://doi.org/10.3748/wjg.v20.i18.5442>.
2. Pinzani, Massimo, Matteo Rosselli, and Michele Zuckermann. "Liver Cirrhosis." *Best Practice & Research Clinical Gastroenterology* 25, no. 2 (April 2011): 281–90. <https://doi.org/10.1016/j.bpg.2011.02.009>.
3. Muir, Andrew J. "Understanding the Complexities of Cirrhosis." *Clinical Therapeutics* 37, no. 8 (August 2015): 1822–36. <https://doi.org/10.1016/j.clinthera.2015.05.507>.
4. Wang, Zhenpeng, Aimin Zhang, Yue Yin, Jiashu Tian, Xialin Wang, Zhihong Yue, Lin Pei, et al. "Clinical Prediction of HBV-Associated Cirrhosis Using Machine Learning Based on Platelet and Bile Acids." *Clinica Chimica Acta* 551 (November 2023): 117589. <https://doi.org/10.1016/j.cca.2023.117589>.
5. Prakash, K., and S. Saradha. "Efficient Prediction and Classification for Cirrhosis Disease Using LBP, GLCM and SVM from MRI Images." *Materials Today: Proceedings* 81 (2023): 383–88. <https://doi.org/10.1016/j.matpr.2021.03.418>.

**How does your proposal relate to existing work on this topic described in these papers? (200 words)**

The referenced papers collectively cover evidence-based treatment approaches, the complexities of cirrhosis, and clinical predictions.

**Paper 1 - Treatment Perspectives:** Emphasizes the importance of developing a predictive model. It aims to enhance the accuracy of cirrhosis analysis, contributing to the evidence-based decision-making process.

**Paper 2 - Advanced Approach:** Provides a comprehensive overview, the project further enhances understanding through predictive modelling, clustering, and interdisciplinary analysis, aiming to contribute to more effective clinical decision-making in cirrhosis management.

**Paper 3 - Understanding the Complexity:** Emphasis on understanding the complexities of cirrhosis. The combination of medical insights with machine learning reflects an understanding, striving for analysis beyond traditional methods.

**Paper 4 - Clinical Prediction Using Machine Learning:** Introduces machine learning in clinical prediction for HBV-associated cirrhosis. This resonates by employing Logistic Regression, Hyperparameter Tuning, Random Forest, and Gradient Boosting for enhanced predictive accuracy.

**Paper 5 - Efficient Prediction Using Imaging:** Prediction from MRI images, it extends the predictive modelling scope by incorporating various clinical and demographic features, illustrating a broader approach to cirrhosis analysis.

In summary, the proposed project extends existing work by incorporating advanced predictive modelling techniques, and making use of insights from medical research, contributing to more understanding of cirrhosis for improved clinical decision-making.

**What are the research questions that you will attempt to answer? (200 words)**

Research questions guides and helps in focusing on the key aspects to be explored.

1. How do transformed features, such as age categories and diagnostic indicators, contribute to the predictive accuracy of cirrhosis models?
2. How can clusters formed through K-medoids analysis contribute to an understanding of cirrhosis subgroups?

**How will you explore these questions?**

**- What software and programming environment will you use?**

Python programming language with popular libraries such as Pandas, NumPy, Scikit-learn, and others in the Jupyter notebook as it offers a rich ecosystem of tools and libraries.

**- What coding/development will you do?**

Implementing data preprocessing, clustering algorithms (K-medoids), and predictive models (Logistic Regression, and Random Forest). Custom coding is necessary to tailor the analysis to the specific requirements of the project to get more accuracy.

**- What data will be used for your investigations?**

A comprehensive dataset containing clinical and demographic features relevant to cirrhosis. A rich dataset with features like Age, Bilirubin, Cholesterol, Albumin, Copper, ALP, AST, Triglycerides, Platelets, and Prothrombin is crucial for building accurate predictive models and identifying meaningful patterns.

**- Is this data currently available, if not, where will it come from?**

The data is available publicly on Kaggle and we can also gather data from different sources on the internet, medical databases or collaborative research efforts.

Datasets with Cirrhosis details:

1. <https://www.kaggle.com/datasets/muki2003/lung-dataset>
2. <https://www.kaggle.com/datasets/fedesoriano/cirrhosis-prediction-dataset>
3. <https://www.kaggle.com/datasets/uciml/indian-liver-patient-records>

Dataset for hospital details:

4. <https://www.kaggle.com/datasets/ezhilnandhini/hospital-details>

**- What experiments do you expect to run?**

Experiments involving clustering with K-medoids and predictive analytics with Logistic Regression, Hyperparameter Tuning, Random Forest, and Gradient Boosting. Each experiment serves a specific purpose in uncovering patterns and optimizing predictive accuracy.

**- What output do you expect to gather?**

Output will include predictive model results, cluster assignments, and insights into feature importance. The output provides actionable information for clinical decision-making and a deeper understanding of cirrhosis.

**- How will the results be evaluated?**

Employ evaluation metrics such as accuracy, precision, recall, and F1 score for predictive models. Use cluster validity indices for K-medoids. Robust evaluation metrics ensure the reliability and effectiveness of the developed models and clusters.