SCHOOL OF BIOTECHNOLOGY AND BIOINFORMATICS
D. Y. PATIL DEEMED TO BE UNIVERSITY, NAVI MUMBAI
Plot No. 50, Sector 15, CBD Belapur, Navi Mumbai – 400 614.

DIPLOMA IN DATA SCIENCE FOR HEALTHCARE
SYLLABUS

In academic collaboration with

BRAASK EDUCATION PVT. LTD,

Title of the Programme

Diploma in Data Science for Healthcare

About the Programme

The Diploma in Data Science for Healthcare is a 10-month, 40-credit, project-based online learning programme designed in alignment with the principles of the National Education Policy (NEP) 2020, offered by the School of Biotechnology and Bioinformatics, D.Y. Patil Deemed to be University, Navi Mumbai, in academic collaboration with TechMedBuddy. Designed at the intersection of medical Informatics, technology, and innovation, this program is tailored to meet the growing demand for skilled professionals in the rapidly evolving landscape of digital healthcare.

This interdisciplinary program blends foundational medical and life sciences knowledge with cutting-edge advancements in data science, computational biology, and digital health technologies. Through a carefully curated curriculum and hands-on learning approach, participants will gain deep insights into the application of Artificial Intelligence (AI), machine learning, and data-driven decision-making in the healthcare ecosystem.

Whether you’re a healthcare professional aiming to upskill, a life sciences graduate looking to transition into digital health, or a tech enthusiast passionate about healthcare innovation, this program empowers you to become a leader in the future of AI-driven, patient-centric care.

Graduates of this program will be well-equipped to take on roles in clinical informatics, health tech product development, research and development, and healthcare analytics across hospitals, biotech companies, startups, and global healthcare organisations.

Programme Objectives

The primary objectives of the programme are to:

- Provide interdisciplinary knowledge integrating healthcare and data science.
- Develop proficiency in analytical tools and programming languages used in biomedical data analysis.
- Equip students with skills to handle, analyse, and interpret complex healthcare datasets.
- Foster an understanding of ethical and regulatory aspects of digital health solutions.
- Prepare learners for careers in research, healthtech, pharma, and clinical informatics.

Learning Outcomes

Upon successful completion of the programme, students will be able to:

- Apply statistical, computational, and machine learning techniques to healthcare data.
- Analyse omics, clinical, and imaging datasets for biomedical applications.
- Design and implement AI-based solutions for healthcare challenges.
- Navigate regulatory frameworks and ethical considerations in health data science.
- Contribute to the development and deployment of digital health technologies and products.

Eligibility Criteria

The programme is open to a broad spectrum of learners from diverse academic and professional backgrounds who are keen to build expertise in healthcare artificial intelligence and biomedical data science. Eligible candidates include:

- Students who have completed a minimum of 2 Years of undergraduate studies from a recognised University in Medical Sciences, Dental Sciences, Pharmacy, Nursing, Life Sciences, Biomedical Science, Biotechnology, Bioinformatics, Microbiology, Molecular Biology, any other life sciences-related discipline and Computer Science and IT.
- Postgraduate students currently pursuing or having completed a Master's degree in Medical Sciences, Dental Sciences, Pharmacy, Nursing, Life Sciences, Biomedical Science, Biotechnology, Bioinformatics, Microbiology, Molecular Biology, any other life sciences-related discipline and Computer Science and IT.
- PhD scholars and postdoctoral researchers seeking interdisciplinary exposure in healthcare data analytics and AI.
- Working professionals in the healthcare, pharmaceutical, research, or technology sectors who are looking to upskill or transition into health data science roles.

Applicants should demonstrate a strong interest in interdisciplinary research, data analysis, and the application of artificial intelligence in healthcare and life sciences. A foundational understanding of biology and/or computational methods will be beneficial, but is not mandatory.

Course Duration and Credit Structure

This is a 10-month, **40-credit, project-based online Diploma programme** comprising more than 600 hours of live, interactive lectures and practical sessions, designed in compliance with NEP2020.

Career Pathways

Graduates of the programme will be prepared for diverse roles in the following domains:

- Health Data Analyst / Clinical Data Scientist
- Bioinformatics Specialist
- AI/ML Engineer in Healthcare
- Digital Health Product Developer
- Research Associate in Translational Medicine
- Regulatory & Compliance Analyst
- Personalised Medicine Analyst

They will find employment opportunities in hospitals, pharmaceutical companies, digital health startups, research institutes, public health organisations, and government health departments.

Course Selection Policy and Credit Structure

Semester	Type of Course	Requirements	Credits
Semester I	Core / Foundational Courses	PCC - 3 PEC - 2/3	19 credits
Semester II	Elective / Specialisation & Capstone	PCC - 4 PEC/ELEC 1/2	21 credits
		Total	**40 Credits**

List of Abbreviations

Abbreviation	Title
PCC	Program Core Course
PEC	Program Elective Course
ELC	Experiential Learning Course

Program Course/Credit Structure

Semester - I									
Paper Code	Category	Course	Hours per week			Credit			Total Credit
			L	T	P	L	T	P	
2503DDC1T1	**PCC**	Foundations of Data Science in Healthcare	3	2	–	3	2	–	5
2503DDC1T2	**PCC**	Applied Machine Learning for Health	3	2	–	3	2	–	5
2503DDC1T3	**PCC**	Biomedical Data Ecosystem	3	–	–	3	–	–	3
Discipline Specific Elective (Any 2)									
2503DDS1W1	**ELC**	NLP & LLM for Clinical Applications	2	-	2	2	-	1	(2 + 1) + (2 + 1)
2503DDS1W2	**ELC**	Biomedical Imaging & Computer Vision	2	-	2	2	-	1	
2503DDS1W3	**ELC**	Omics & Genomics Data Science	2	-	2	2	-	1	
Total			**13**	**4**	**4**	**15**	**4**	**3**	**19**

Note: Out of 22 Credits, students need to complete 19 Credits.

Semester - II									
Paper Code	**Category**	**Course**	**Hours per week**			**Credit**			**Total Credit**
			L	**T**	**P**	**L**	**T**	**P**	
2503DDC2T1	**PCC**	Generative AI & Innovation in HealthTech	2	1	–	2	**1**	**-**	3
2503DDC2T2	**PCC**	Model Development, MLOps	2	1	–	2	**1**	**–**	3
2503DDC2T3	**PCC**	AI for CADD	2	1	-	2	**1**	**-**	3
2503DDC2T3	**PCC**	Responsible AI, Ethics & Policy in Healthcare	3	–	–	3	**–**	**–**	3
Discipline Specific Elective (Any 1)									
2503DDK2P1	**ELC**	Project 1* / Project 2*	–	–	6	–	**—**	**3**	3
2503DDS2T1	**PEC**	Multi-Modality in Healthcare Applications	2	1	–	2	**1**	**–**	
2503DDS2T2	**PEC**	Biomedical Time Series & Sensor Data	2	1	–	2	**1**	**-**	
Mandatory Capstone Project									
2503DDK2P2	**PEC/ELC**	Capstone Project **	–	–	12	–	**–**	**6**	6
Total			**13**	**5**	**27**	**13**	**5**	**9**	**21**

Note:

* Project 1 (from Semester 1) and Project 2 (from Semester 2) are independent projects. Students can choose any one course from the respective semester and complete a project based on it. Each project includes 6 hours of guided learning, along with dedicated mentorship and support throughout the process.

** This is a mandatory Capstone Project that carries 6 credits. Students will receive 12 hours of dedicated mentorship and support to guide them through the project.

Evaluation Scheme

Diploma in Data Science for Healthcare						
Paper Code	**Course**	**Internal Assessment**	**End Semester Examination**	**Total**	**Credit**	**Sem**
2503DDC1T1	Foundations of Data Science in Healthcare	40	60	100	5	I
2503DDC1T2	Applied Machine Learning for Health	40	60	100	5	I
2503DDC1T3	Biomedical Data Ecosystem	40	60	100	3	I
2503DDS1W1	NLP & LLM for Clinical Applications	40	60	100	3	I
2503DDS1W2	Biomedical Imaging & Computer Vision	40	60	100	3	I
2503DDS1W3	Omics & Genomics Data Science	40	60	100	3	I
2503DDC2T1	Generative AI & Innovation in HealthTech	40	60	100	3	II
2503DDC2T2	Model Development, MLOps	40	60	100	3	II
2503DDC2T3	AI for CADD	40	60	100	3	II
2503DDC2T3	Responsible AI, Ethics & Policy in Healthcare	40	60	100	3	II
2503DDK2P1	Project 1* OR Project 2*	Complete Independent Project		100	3	II
2503DDS2T1	Multi-Modality in Healthcare Applications	40	60	100	3	II
2503DDS2T2	Biomedical Time Series & Sensor Data	40	60	100	3	II
2503DDK2P2	Capstone Project **	Complete Independent Project			6	

Employable Skills

After completing the Diploma, the student:

1. Qualifies for key industry roles such as Clinical Data Scientist, Bioinformatics Analyst, AI/ML Engineer (Healthcare), NLP Scientist (Clinical), and Genomics Data Scientist in healthtech, biotech, and pharmaceutical companies.

2. Demonstrates strong programming skills in Python, R, SQL, and MATLAB, with practical experience using domain-specific libraries and tools for biomedical data science.

3. Applies advanced machine learning and deep learning techniques (e.g., CNNs, RNNs, Transformers) to solve problems in clinical diagnostics, personalized medicine, and drug discovery.

4. Works confidently with multi-omics, medical imaging, time-series, and clinical text data, enabling real-world contributions to product pipelines in digital health and precision medicine.

5. Builds, deploys, and monitors end-to-end AI models using industry tools such as Git, Docker, MLflow, FastAPI, and Streamlit, following modern MLOps workflows.

6. Understands healthcare regulations and ethical AI practices, including HIPAA, GDPR, clinical safety, and algorithmic fairness — essential for deployment in regulated environments.

7. Communicates data-driven insights effectively, using tools like Tableau and Power BI to support clinical, product, and research decision-making across interdisciplinary teams.

8. Aligns with hiring needs of companies like Qure.ai, Invitae, Strand Life Sciences, Practo, Tempus, Novartis, and digital health arms of Google, Microsoft, and AWS, while also being prepared for research or doctoral studies globally.

Semester – 1

2503DDC1T1	**Foundations of Data Science in Healthcare**	**Credits (5)**	**Hours (75)**
COURSE OBJECTIVES: TO ENABLE THE STUDENTS TO-			
1	To build a strong foundation in mathematical, statistical, and computational concepts essential for biomedical data analysis.		
2	To familiarize learners with the principles of data acquisition, preprocessing, and quality control in biological and clinical research.		
3	To equip participants with programming skills in R and Python and tools like SQL and Git for efficient data handling and reproducible research.		
4	To introduce visualization techniques and experimental design for effective communication and robust analysis in biomedical sciences.		
5	To develop the ability to apply inferential statistics and causal inference methods to derive meaningful insights from real-world biomedical datasets.		
COURSE OUTCOMES: THE STUDENTS WILL BE ABLE TO-			
1	Apply elementary mathematical and linear algebra concepts to interpret high-dimensional biological data.		
2	Design, clean, and preprocess biomedical datasets using appropriate statistical and computational methods.		
3	Write basic to intermediate-level code in Python and R for data exploration, manipulation, and visualization.		
4	Execute and interpret inferential statistical analyses and causal models relevant to biomedical research.		
5	Utilize modern data tools such as SQL for querying databases, Git for version control, and Tableau/Power BI for visual analytics.		

COURSE CONTENT

MODULE NO.1	**Quantitative Foundations for Biologists**	**7**

- Functions and equations
- Logarithms and exponentials in biological modeling
- Basic probability concepts
- Growth models (linear vs. exponential)
- Introduction to combinatorics and factorial-based problems

MODULE NO.2	**Linear Algebra for Data Science Applications**	**10**

- Vectors, matrices, and operations
- Matrix multiplication and linear transformations
- Eigenvalues and eigenvectors
- Orthogonality, projections, and SVD
- Applications in PCA and machine learning

MODULE NO.3	Statistical Exploration, Experimental Design, and Visualization	10

- Measures of central tendency and dispersion
- Experimental vs observational study design
- Randomization, control groups, bias
- Data visualization principles using ggplot2/Matplotlib
- Effective storytelling with data

MODULE NO.4	Inferential Statistics in Biomedical Research	10

- Confidence intervals and hypothesis testing
- t-tests, chi-square, ANOVA
- Non Parametric Tests
- Statistical power and Type I/II errors
- Application to genomics and clinical trial data

MODULE NO.5	Programming Essentials for Data Science	20

- Python & R basics: variables, loops, conditionals
- Functions, libraries (pandas, tidyverse)
- File I/O, data wrangling, string manipulations
- Basic debugging and scripting

MODULE NO.6	Advanced Statistical Inference and Causal Modelling	10

- Revisiting hypothesis testing frameworks
- Introduction to causal diagrams (DAGs)
- Confounding, mediation, and adjustment
- Instrumental variables and matching methods
- Potential outcomes framework

MODULE NO.7	Collaborative Coding and Version Control	3

- Git basics: init, clone, add, commit, push
- Branching and merging workflows
- Resolving conflicts
- Versioning scripts and notebooks
- GitHub/GitLab for collaboration

MODULE NO.8	Relational Databases and SQL for Biomedical Data	5

- Introduction to relational databases
- SELECT, WHERE, GROUP BY, JOIN operations
- Filtering, sorting, aggregations
- Biomedical data querying examples (e.g., patient records, expression tables)
- Using SQL in Python/R pipelines

REFERENCES

1	Strang, G. (2009), *Linear Algebra and Its Applications*, 4th Edition, Cengage Learning, Boston.
2	Ross, S. (2010), *A First Course in Probability*, 8th Edition, Pearson Education, Boston.
3	Meyer, C. D. (2000), *Matrix Analysis and Applied Linear Algebra*, SIAM, Philadelphia

4	Wasserman, L. (2004), *All of Statistics*, Springer, New York.
5	Bruce, P., & Bruce, A. (2017), *Practical Statistics for Data Scientists*, O'Reilly Media, Sebastopol.
6	Szklo, M., & Nieto, F. J. (2014), *Epidemiology: Beyond the Basics*, 3rd Edition, Jones & Bartlett Learning, Burlington.
7	Tufte, E. R. (2001), *The Visual Display of Quantitative Information*, 2nd Edition, Graphics Press, Cheshire.
8	Wayne W. Daniel | Chad L Cross, Biostatistics , A foundation for Analysis in the Health sciences

2503DDC1T2	Applied Machine Learning for Health		Credits (5)	Hours (75)
COURSE OBJECTIVES: TO ENABLE THE STUDENTS TO-				
1	To develop a foundational and applied understanding of machine learning concepts tailored for healthcare data.			
2	To enable learners to implement supervised and unsupervised learning algorithms using real-world biomedical datasets.			
3	To introduce ensemble methods and deep learning architectures relevant to health informatics and clinical prediction.			
4	To provide a strong foundation in model evaluation, validation strategies, and performance metrics.			
5	To instill the importance of model interpretability and ethical AI in healthcare contexts.			
COURSE OUTCOMES: THE STUDENTS WILL BE ABLE TO-				
1	To instill the importance of model interpretability and ethical AI in healthcare contexts.			
2	Apply regression, classification, clustering, and dimensionality reduction methods to biomedical problems. LOINC for structured interoperability.			
3	Develop and train deep learning models (MLPs, CNNs, RNNs) using modern frameworks.			
4	Evaluate model performance using appropriate metrics and validation techniques.			
5	Interpret complex models using SHAP, LIME, and apply best practices in ethical machine learning.			
COURSE CONTENT				
MODULE NO.1	**Foundations of Machine Learning in Healthcare**			**9**
● Introduction to ML in biomedical domains ● Types of machine learning (supervised, unsupervised, reinforcement) ● ML pipeline: preprocessing to deployment ● Role of ML in genomics, EHR, imaging, and epidemiology				
MODULE NO.2	**Supervised Learning – Regression Techniques**			**8**
● Linear and logistic regression ● Regularization (L1, L2) ● Model diagnostics (residuals, RMSE, AUC) ● Case studies in biomarker prediction and lab result estimation				
MODULE NO.3	Supervised Learning – Classification Algorithms			**12**
● k-NN, Decision Trees, Naive Bayes, SVM ● Bias-variance tradeoff ● Multiclass vs multilabel classification ● Applications: disease classification, image diagnosis, risk stratification				
MODULE NO.4	**Unsupervised Learning and Dimensionality Reduction**			**10**

● k-means, hierarchical clustering ● Principal Component Analysis (PCA), t-SNE ● Feature extraction and latent variable discovery ● Applications in cancer subtyping, microbiome, transcriptomics		
MODULE NO.5	**Ensemble Methods for Robust Prediction**	**10**
● Random Forests and Gradient Boosting Machines (XGBoost, LightGBM) ● Bagging vs Boosting ● Feature importance analysis ● Use cases: prognosis modeling, imaging diagnostics		
MODULE NO.6	**Introduction to Deep Learning for Biomedical Applications**	**10**
● Neural networks (MLPs, CNNs, RNNs) ● Activation functions, optimizers, and loss functions ● CNN for imaging data, RNN for sequential health data (e.g., vitals, EHR) ● Intro to frameworks: TensorFlow, Keras, PyTorch		
MODULE NO.7	**Model Evaluation and Validation Strategies**	**8**
● Holdout, k-fold, stratified CV ● Evaluation metrics (Precision, Recall, F1, ROC-AUC, PR Curve) ● Overfitting, underfitting, and hyperparameter tuning ● Cross-validation in imbalanced datasets (SMOTE, weighting)		
MODULE NO.8	**Interpretability and Explainability in ML Models**	**8**
● Black-box vs glass-box models ● SHAP, LIME, Partial Dependence Plots ● Feature attribution in clinical models ● Ethical AI in health: bias detection, fairness		

REFERENCES

1	Hastie, T., Tibshirani, R., & Friedman, J. (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Edition, Springer, New York.
2	Bishop, C. M. (2006), *Pattern Recognition and Machine Learning*, Springer, New York.
3	Breiman, L. (2001), *Statistical Modeling: The Two Cultures*, *Statistical Science*, 16(3), 199-231. DOI: 10.1214/ss/1009213726
4	Goodfellow, Bengio, Courville (2017), DeepLearning, MIT Press.
5	Harrison, Conrad & Sidey-Gibbons, Chris. (2021). Machine learning in medicine: a practical introduction to natural language processing. BMC Medical Research Methodology. 21. 10.1186/s12874-021-01347-1.
6	Habehh, Hafsa & Gohel, Suril. (2021). Machine Learning In Healthcare. Current Genomics. 22. 10.2174/1389202922666210705124359.

2503DDC1T3	Biomedical Data Ecosystem	Credits (3)	Hours (45)
COURSE OBJECTIVES: TO ENABLE THE STUDENTS TO-			
1	Introduce students to the fundamental types, formats, and sources of biomedical data, including clinical, omics, imaging, and sensor-based modalities.		
2	Familiarize learners with key health informatics systems and medical coding standards such as HL7, FHIR, ICD, and LOINC for structured data representation and interoperability.		
3	Equip participants with practical skills in data cleaning, preprocessing, and quality assessment specifically tailored to complex healthcare datasets.		
4	Enable the development of interactive and interpretable visualizations using modern BI tools like Tableau and Power BI for healthcare analytics.		
5	Promote integrative thinking through real-world case studies, focusing on the synthesis of multi-modal biomedical data for clinical decision-making and research insights.		
COURSE OUTCOMES: THE STUDENTS WILL BE ABLE TO-			
1	Identify and classify biomedical data types including omics, clinical, imaging, and sensor data.		
2	Apply health data standards such as HL7, FHIR, ICD, and LOINC for structured interoperability.		
3	Perform data preprocessing and quality assessment tailored to biomedical datasets.		
4	Integrate and analyze multi-modal health data to derive actionable insights in real-world scenarios.		
5	Create insightful dashboards using Tableau and Power BI for clinical and research applications.		

COURSE CONTENT

MODULE NO.1	**Biomedical Data Modalities and Structures**	**6**

- Introduction to structured, semi-structured, and unstructured biomedical data
- Data types: clinical records, multi-omics (genomics, transcriptomics, proteomics, metabolomics), imaging (DICOM), and wearable sensor data
- Data characteristics: volume, velocity, veracity, and variety
- Storage formats and annotation standards

MODULE NO.2	**Electronic Health Record Systems and Medical Coding Frameworks**	**5**

- Architecture of EHR systems
- Interoperability protocols: HL7 v2/v3, FHIR
- Clinical ontologies and vocabularies: SNOMED CT, ICD-10, LOINC, RxNorm
- Integration of genomic and clinical records in EHRs

MODULE NO.3	Interoperability Standards in Digital Health Ecosystems	**5**

- Principles of semantic and syntactic interoperability
- OpenEHR and FHIR-based APIs
- Standardized messaging and exchange protocols
- Global initiatives: OHDSI, GA4GH, and EU/US interoperability efforts

MODULE NO.4	**Biomedical Data Quality, Curation, and Preprocessing Pipelines**	**9**

● Challenges in biomedical data heterogeneity ● Missing data mechanisms (MCAR, MAR, MNAR) ● Data normalization, outlier detection, deduplication ● Annotation validation and feature harmonization strategies		
MODULE NO.5	**Advanced Visualization Techniques for Biomedical Informatics**	**10**
● Visual analytics using clinical and multi-omics datasets ● Interactive dashboards using Tableau and Power BI ● KPI metrics for healthcare outcomes ● Mapping EHR data to time-series plots and heatmaps		
MODULE NO.6	**Real-World Case Studies in Biomedical Data Integration**	**10**
● Integrated data analysis from multi-modal sources (EHR + Omics + Imaging) ● Use cases: cancer subtype prediction, clinical trial matching, risk stratification ● Hands-on exploration of real-world datasets: TCGA, MIMIC-III, UK Biobank ● Ethical considerations, privacy, and data governance		

REFERENCES

1	Edward H., JamesJ.Cimino. (2018), Biomedical Informatics: Computer Applications in Healthcare and Biomedicine, 4th Edition, Springer, New York.
2	Gordon D. Brown, Timothy B. Patrick, Kalyan S. Pasupathy (2012), *Health Informatics: A Systems Perspective*, 2nd Edition, Springer, New York.
3	Milligan, J. N. (2019), *Learning Tableau 2019 - Third Edition*, Packt Publishing, Birmingham.
4	Mathé, E., Hays, J. L., Stover, D. G., & Chen, J. L. (2018), *The Omics Revolution Continues: The Maturation of High-Throughput Biological Data Sources*, *Nature Reviews Genetics*, 19(11), 745-758. DOI: 10.1038/s41576-018-0034-9.
5	Hall JL, Ryan JJ, Bray BE, et al. Merging Electronic Health Record Data and Genomics for Cardiovascular Research: A Science Advisory From the American Heart Association. *Circ Cardiovasc Genet*. 2016;9(2):193-202. doi:10.1161/HCG.0000000000000029

<table>
<tr><th colspan="2">2503DDS1W1</th><th>NLP and LLM for Clinical Applications</th><th>Credits (3)</th><th>Hours (45)</th></tr>
<tr><td colspan="5">COURSE OBJECTIVES: TO ENABLE THE STUDENTS TO-</td></tr>
<tr><td>1</td><td colspan="2">To introduce core NLP techniques tailored for structured and unstructured clinical text data.</td><td></td><td></td></tr>
<tr><td>2</td><td colspan="2">To familiarize students with word embedding models and biomedical language representations.</td><td></td><td></td></tr>
<tr><td>3</td><td colspan="2">To explore transformer and generative models (BERT, GPT, BioBERT) for clinical tasks.</td><td></td><td></td></tr>
<tr><td>4</td><td colspan="2">To apply large language models to summarize, extract, and generate clinical knowledge.</td><td></td><td></td></tr>
<tr><td>5</td><td colspan="2">To understand the ethical, interpretability, and integration challenges in deploying NLP for healthcare.</td><td></td><td></td></tr>
<tr><td colspan="5">COURSE OUTCOMES: THE STUDENTS WILL BE ABLE TO-</td></tr>
<tr><td>1</td><td colspan="2">Process and extract information from raw clinical text using NLP pipelines and NER tools.</td><td></td><td></td></tr>
<tr><td>2</td><td colspan="2">Use domain-specific word embeddings and contextual models for downstream tasks.</td><td></td><td></td></tr>
<tr><td>3</td><td colspan="2">Apply BERT-based and GPT-based models to solve clinical NLP problems.</td><td></td><td></td></tr>
<tr><td>4</td><td colspan="2">Build retrieval-augmented generation systems for evidence-aware text generation.</td><td></td><td></td></tr>
<tr><td>5</td><td colspan="2">Evaluate, interpret, and ethically deploy NLP models within clinical applications.</td><td></td><td></td></tr>
<tr><td colspan="5">COURSE CONTENT</td></tr>
<tr><td colspan="2">MODULE NO.1</td><td colspan="2">Foundations of Clinical Text Analytics</td><td>8</td></tr>
<tr><td colspan="5">● Introduction to Natural Language Processing in healthcare
● Clinical narratives, discharge summaries, pathology reports
● Text normalization, stemming, lemmatization
● Named Entity Recognition (NER) for drugs, diagnoses, symptoms
● Tools: spaCy, scispaCy, cTAKES</td></tr>
<tr><td colspan="2">MODULE NO.2</td><td colspan="2">Biomedical Word Embeddings and Contextual Representations</td><td>8</td></tr>
<tr><td colspan="5">● Bag-of-Words vs TF-IDF vs dense vector representations
● Word2Vec, GloVe: principles and training
● Domain-specific embeddings: BioWordVec, ClinicalVec, PubMed embeddings
● Use cases in information extraction and clustering</td></tr>
<tr><td colspan="2">MODULE NO.3</td><td colspan="2">Transformer-based Architectures for Clinical NLP</td><td>8</td></tr>
<tr><td colspan="5">● Anatomy of the Transformer: attention mechanisms and encoders
● BERT and fine-tuning strategies
● Biomedical variants: BioBERT, ClinicalBERT, BlueBERT
● Text classification and sequence labeling in EMR and radiology data</td></tr>
<tr><td colspan="2">MODULE NO.4</td><td colspan="2">Generative Language Models in Clinical Workflows</td><td>8</td></tr>
<tr><td colspan="5">● Introduction to GPT models: GPT-2, GPT-3, GPT-4
● Fine-tuning vs prompt engineering in clinical tasks
● Text summarization: discharge summaries, radiology reports
● Clinical question answering systems using generative models</td></tr>
</table>

MODULE NO.5	Retrieval-Augmented Generation (RAG) and Knowledge-Enhanced NLP	8

- Hybrid architecture of retrieval + generation
- FAISS for similarity search, vector stores
- Connecting unstructured EHRs with structured biomedical knowledge graphs
- Applications in drug discovery, evidence-based reasoning

MODULE NO.6	Ethics, Interpretability in Clinical NLP	5

- Bias in medical language models
- Model explainability (SHAP, attention maps)
- De-identification of protected health information (PHI)

REFERENCES

1	Jurafsky, D., & Martin, J. H. (2020), *Speech and Language Processing*, 3rd Edition, Draft available online. Access the draft here.
2	Allen, J. (1995), *Natural Language Understanding*, 2nd Edition, Benjamin/Cummings, Menlo Park.
3	Mitkov, R. (Ed.) (2003), *The Oxford Handbook of Computational Linguistics*, 1st Edition, Oxford University Press, Oxford.
4	Lee, J., Yoon, W., Kim, S., & Kim, D. (2020), *BioBERT: A Pre-trained Biomedical Language Representation Model for Biomedical Text Mining*, *Bioinformatics*, 36(4), 1234-1240. DOI: 10.1093/bioinformatics/btz682
5	Alsentzer, E., et al. (2019), *Publicly Available Clinical BERT Embeddings (ClinicalBERT)*, *arXiv preprint*, arXiv:1904.03323. DOI: 10.48550/arXiv.1904.03323
6	Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018), *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*, *arXiv preprint*, arXiv:1810.04805. DOI: 10.48550/arXiv.1810.04805

2503DDS1W2	Biomedical Imaging and Computer Vision		Credits (3)	Hours (45)
COURSE OBJECTIVES: TO ENABLE THE STUDENTS TO-				
1	To introduce medical imaging modalities and their applications in diagnostic workflows.			
2	To provide foundational understanding of image preprocessing and classical CV techniques in biomedical data.			
3	To train learners in state-of-the-art deep learning models for object detection and segmentation.			
4	To develop the ability to construct full pipelines from raw images to classification and diagnosis.			
5	To explore domain-specific applications of computer vision in radiology, pathology, and clinical diagnostics.			
COURSE OUTCOMES: THE STUDENTS WILL BE ABLE TO-				
1	Understand various medical imaging modalities and preprocess biomedical image data.			
2	Apply traditional and deep learning-based CV techniques to extract features and identify patterns.			
3	Implement object detection models such as YOLO and Faster-RCNN for identifying key anatomical regions.			
4	Train segmentation models (U-Net, Mask-RCNN) for accurate delineation of biomedical structures.			
5	Deploy CV pipelines to automate diagnostic workflows in pathology and radiology.			

COURSE CONTENT

MODULE NO.1	Fundamentals of Biomedical Imaging Modalities	6

- Overview of imaging technologies: MRI, CT, PET, X-ray, Ultrasound
- Histopathological imaging and whole slide images (WSIs)
- Resolution, contrast, and modality-specific challenges
- DICOM standards and PACS integration

MODULE NO.2	Image Preprocessing and Augmentation for Medical Data	10

- Noise reduction, histogram equalization, normalization
- Resizing, cropping, rotation, and flipping
- CLAHE, adaptive thresholding, and color space transformations
- Data augmentation strategies in medical imaging

MODULE NO.3	Feature Extraction and Traditional Computer Vision Methods	5

- Edge detection (Sobel, Canny), contour detection
- Texture analysis (Haralick features, LBP)
- Thresholding, morphological operations
- Region of interest (ROI) extraction and shape-based analysis

MODULE NO.4	Object Detection in Medical Imaging	8
● Object localization vs classification ● YOLO (You Only Look Once): architecture and use cases ● Faster-RCNN and region proposal networks ● Use cases in tumor detection and organ localization		
MODULE NO.5	**Semantic and Instance Segmentation Techniques**	**8**
● Image segmentation challenges in biomedical data ● U-Net: architecture, skip connections, variations (ResU-Net, Attention U-Net) ● Mask-RCNN for instance-level segmentation ● Evaluation metrics: Dice score, Jaccard index, Hausdorff distance		
MODULE NO.6	**Deep Learning Pipelines in Radiology and Pathology**	**8**
● Pre-trained models (ImageNet, MedMNIST, RadImageNet) ● Transfer learning and domain adaptation ● Multi-class classification of histopathological slides ● Deep learning for radiology report generation		

REFERENCES	
1	Prince, J. L., & Links, J. (2006), *Medical Imaging Signals and Systems*, Pearson Prentice Hall, Upper Saddle River.
2	Zhou, Y., Greenspan, H., & Shen, D. (2018), *Deep Learning for Medical Image Analysis*, Elsevier, Amsterdam.
3	Szeliski, R. (2010), *Computer Vision: Algorithms and Applications*, Springer, New York.
4	Perez, L., & Wang, J. (2017), *The Effectiveness of Data Augmentation in Image Classification using Deep Learning*, *arXiv preprint*, arXiv:1712.04621. DOI: 10.48550/arXiv.1712.04621.
5	Ronneberger, O., Fischer, P., & Brox, T. (2015), *U-Net: Convolutional Networks for Biomedical Image Segmentation*, *MICCAI 2015*, 234-241. DOI: 10.1007/978-3-319-24574-4_28.
6	He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017), *Mask R-CNN*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(2), 296-307. DOI: 10.1109/TPAMI.2017.2699184.

2503DDS1W3	Omics and Genomics Data Science		Credits (3)	Hours (45)
COURSE OBJECTIVES: TO ENABLE THE STUDENTS TO-				
1	To provide a comprehensive understanding of high-throughput omics technologies and their applications in biomedical research.			
2	To train students in quality control, alignment, and normalization workflows using standard bioinformatics tools.			
3	To introduce ensemble methods and deep learning architectures relevant to health informatics and clinical prediction.			
4	To introduce spatial transcriptomics and its integration with classical omics layers.			
5	To empower learners to perform multi-omics data integration and identify molecular signatures for disease stratification and prognosis.			
COURSE OUTCOMES: THE STUDENTS WILL BE ABLE TO-				
1	Distinguish between various genomic and transcriptomic technologies and choose appropriate platforms for research questions.			
2	Process and analyze bulk and single-cell transcriptomics datasets using modern bioinformatics pipelines.			
3	Perform quality control, normalization, and alignment of omics data with proficiency.			
4	Explore spatial transcriptomics data and interpret spatial expression patterns.			

COURSE CONTENT

MODULE NO.1	Principles of Genomic Technologies and High-throughput Platforms	4

- Central dogma and evolution of omics technologies
- Overview of NGS, microarrays, and emerging long-read platforms
- Types of omics: genomics, transcriptomics, epigenomics, proteomics
- Experimental design considerations for omics studies

MODULE NO.2	Transcriptomics – From Bulk RNA-Seq to Microarrays	6

- RNA extraction, library prep, and sequencing strategies
- Bulk RNA-Seq vs. Microarrays: strengths and limitations
- Data formats (FASTQ, CEL, count matrices)
- Introduction to analysis workflows (DESeq2, Limma)

MODULE NO.3	Single-cell Transcriptomics and Cellular Heterogeneity	10

- Principles of single-cell RNA-seq (10x Genomics, Smart-seq2)
- scRNA-seq preprocessing (cell filtering, normalization)
- Dimensionality reduction (PCA, UMAP, t-SNE)
- Cell clustering, marker gene identification, trajectory analysis

MODULE NO.4	Quality Control, Alignment, and Normalization Techniques	10

- FASTQ QC (FastQC), adapter trimming (Trimmomatic)
- Read alignment (STAR, HISAT2, BWA)

● Transcript quantification (Salmon, Kallisto) ● Batch correction and normalization (TPM, RPKM, TMM, log-transformation)		
MODULE NO.5	**Spatial Transcriptomics and Emerging Omics Frontiers**	**8**
● Introduction to spatially resolved transcriptomics (Visium, MERFISH) ● Integration of spatial and scRNA-seq data ● Tissue architecture inference and cell-type localization ● Future directions: spatial proteomics, in situ sequencing		
MODULE NO.6	**Multi-Omics Integration and Signature Discovery**	**7**
● Integrating transcriptomics, proteomics, and methylation data ● Introduction to mixOmics, MOFA, and DIABLO ● Biomarker and signature discovery ● Case studies: cancer subtyping, treatment response prediction		

REFERENCES	
1	Brown, T. A. (2006), *Genomes 4*, Garland Science, New York.
2	Pevsner, J. (2015), *Bioinformatics and Functional Genomics*, 3rd Edition, Wiley-Blackwell, New Jersey.
3	Mardis, E. R. (2008), *Next-generation DNA sequencing methods*, *Nature Reviews Genetics*, 9(7), 365-376. DOI: 10.1038/nrg2481.
4	Shendure, J., et al. (2017), *DNA sequencing at 40: Past, present and future*, *Nature*, 550(7676), 345-353. DOI: 10.1038/nature24286.
5	Meyerson, M., Gabriel, S., & Getz, G. (2010). Advances in understanding cancer genomes through second-generation sequencing. *Nature reviews. Genetics*, *11*(10), 685–696. https://doi.org/10.1038/nrg2841

2503DDC2T1	Generative AI and Innovation in HealthTech		Credits (3)	Hours (45)
COURSE OBJECTIVES: TO ENABLE THE STUDENTS TO-				
1	To provide a strong foundation in generative AI techniques and their role in healthcare innovation.			
2	To train learners in generating synthetic data for privacy and rare disease applications.			
3	To explore LLMs and prompt engineering for automating clinical documentation and assistance.			
4	To guide students in designing and prototyping GenAI-driven digital health solutions.			
5	To understand the emerging landscape of agentic AI systems and ethical implications in HealthTech.			
COURSE OUTCOMES: THE STUDENTS WILL BE ABLE TO-				
1	Apply generative models (GANs, diffusion) to synthesize realistic biomedical data.			
2	Create privacy-preserving and fair synthetic datasets for clinical research.			
3	Use LLMs for automating documentation and building health-specific AI assistants.			
4	Prototype AI-first products using GenAI APIs and modern UI/UX frameworks.			
5	Understand and evaluate the use of autonomous, agentic AI systems in next-gen healthcare platforms.			

COURSE CONTENT

MODULE NO.1	Foundations of Generative AI for Product Thinking	5

- Role of GenAI in digital health innovation
- Types of generative models (LLMs, Diffusion, GANs, VAEs)
- API-based GenAI services (OpenAI, Cohere, Hugging Face)
- Ideating product features powered by generative models

MODULE NO.2	Prompt Engineering for Clinical Product Applications	8

- Prompt templates for summarization, rephrasing, report generation
- Chain-of-thought, few-shot, and zero-shot prompting
- Tools: LangChain, PromptLayer, LlamaIndex
- Embedding APIs and context management

MODULE NO.3	Large Language Models for Clinical Automation	8

- LLMs in healthcare: OpenAI GPT, Med-PaLM, LLaMA-Med
- Applications: automated documentation, discharge summary generation
- Fine-tuning vs in-context learning
- Integration with EHR and clinical workflows

MODULE NO.4	LLM-Based Workflows and Assistive Health Products	8

- Automating documentation (SOAP notes, discharge summaries)
- Clinical chatbot assistants: backend logic + conversational flows
- Retrieval-Augmented Generation (RAG) for contextual Q&A
- Integrating LLMs into EMRs and patient-facing tools

MODULE NO.5	Synthetic Data Generation and Privacy-Compliant APIs	8
● Generating synthetic EHR and tabular health data with APIs ● Use cases: training models, simulations, de-identification ● Tools: Syntegra, Gretel AI, SDGym ● Regulatory concerns and ethical synthetic data design		
MODULE NO.6	**Rapid Prototyping Using GenAI Toolkits**	**8**
● Streamlit, Gradio, and Flask for health AI MVPs ● Designing frontend + backend API integration ● Deploying GenAI tools with Hugging Face Spaces or Render ● Feedback loops and product iteration strategies		

REFERENCES	
1	Langr, J., & Bok, V. (2019), *GANs in Action: Deep Learning with Generative Adversarial Networks*, Manning Publications, Shelter Island.
2	Hunter, N. (2023), *The Art of Prompt Engineering with ChatGPT*, Independently Published, New York.
3	Goodfellow, I. J., et al. (2014) – *Generative Adversarial Nets*
4	Radford, A., Metz, L., & Chintala, S. (2016) – *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks (DCGANs)*
5	Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017) – *Progressive Growing of GANs for Improved Quality, Stability, and Variation*
6	Brock, A., Donahue, J., & Simonyan, K. (2019) – *Large Scale GAN Training for High Fidelity Natural Image Synthesis*

2503DDC2T2	Model Development and Deployment, MLOps		Credits (3)	Hours (45)
COURSE OBJECTIVES: TO ENABLE THE STUDENTS TO-				
1	To enable students to design, build, and deploy machine learning models in a structured, industry-standard workflow.			
2	To introduce tools like Docker, Flask/FastAPI for efficient ML API deployment.			
3	To equip learners with foundational MLOps practices for version control, monitoring, and reproducibility.			
4	To provide hands-on experience in dashboard development using modern visualization frameworks.			
5	To simulate real-world experience through a healthcare-focused capstone project covering the complete AI product lifecycle.			
COURSE OUTCOMES: THE STUDENTS WILL BE ABLE TO-				
1	Develop and evaluate end-to-end ML models with reproducible pipelines.			
2	Build and deploy ML APIs using Flask/FastAPI integrated with Docker and CI/CD.			
3	Implement MLOps best practices using MLflow, DVC, and monitoring tools.			
4	Create interactive dashboards for clinical and predictive AI applications.			
5	Deliver a complete healthcare AI solution in a capstone project, demonstrating full-cycle model deployment skills.			

COURSE CONTENT

MODULE NO.1	End-to-End ML Model Development Workflow	5

- Problem scoping, data exploration, and preprocessing
- Feature engineering and data pipeline design
- Model selection, tuning, and cross-validation
- Model packaging and reproducibility

MODULE NO.2	Building and Deploying APIs for ML Models	8

- Introduction to Flask and FastAPI for serving ML models
- API design: routing, schema validation, error handling
- Creating REST endpoints for predictions
- Securing and testing APIs

MODULE NO.3	Containerization and Continuous Integration	8

- Docker fundamentals: Dockerfile, image creation, container management
- Docker Compose for multi-container apps
- CI/CD pipelines using GitHub Actions or GitLab CI
- Deployment to cloud platforms (AWS, GCP, Azure – overview)

MODULE NO.4	MLOps Essentials for Scalable AI Systems	8

- Introduction to MLOps principles and lifecycle
- Model versioning using MLflow and DVC

● Logging, monitoring, and automated model retraining ● Experiment tracking and reproducibility in production workflows		
MODULE NO.5	**Interactive Dashboards and Frontend Integration**	**8**
● Streamlit and Gradio for model visualization ● Building dashboards with real-time interaction ● User input forms and backend integration ● Sharing, hosting, and feedback collection		
MODULE NO.6	**Capstone Project – Real-World Healthcare AI Application**	**8**
● End-to-end AI solution development on a real healthcare dataset ● Model training, evaluation, deployment, and dashboarding ● Documentation and report generation ● Peer review and final presentation		

REFERENCES

1	Huyen, C. (2022), *Designing Machine Learning Systems: An Overview for Practical Implementation*, O'Reilly Media, Sebastopol.
2	Poulton, N. (2020), *Docker Deep Dive: A Comprehensive Guide to Docker*, 4th Edition, Leanpub, London.
3	Murray, D. G., & O’Neill, M. P. (2021). *Studying Software Engineering Patterns for Designing Machine Learning Systems*. *arXiv preprint*, arXiv:1910.04736.
4	Kumar, P., & Sharma, V. (2020). *A Study, Analysis, and Deep Dive on Docker Security Vulnerabilities and Their Performance Issues*. *ResearchGate*.
5	Zhang, L., & Li, R. (2021). *A Deep Dive Into How Docker Really Works*. *GitConnected*.
6	Docker, Inc. (2021). *Image Deep-Dive and Best Practices*. *Docker Resources*.

2503DDC2T3	AI for CADD (Computer Aided Drug Discovery)	Credits (3)	Hours (45)
COURSE OBJECTIVES: TO ENABLE THE STUDENTS TO-			
1	To provide a foundational understanding of drug discovery workflows and the role of CADD.		
2	To train students in structure-based and ligand-based design methodologies.		
3	To introduce QSAR modeling techniques using traditional and machine learning approaches.		
4	To explore deep learning frameworks for compound generation, property prediction, and target interaction.		
5	To equip learners with practical skills to implement AI-augmented CADD pipelines.		
COURSE OUTCOMES: THE STUDENTS WILL BE ABLE TO-			
1	Understand the core concepts of drug discovery and differentiate between SBDD and LBDD strategies.		
2	Apply molecular docking and virtual screening to identify lead compounds.		
3	Develop and evaluate QSAR models using machine learning techniques.		
4	Implement deep learning algorithms for drug-likeness prediction and compound generation.		
5	Design and interpret AI-based CADD pipelines using real-world datasets and tools.		

COURSE CONTENT

MODULE NO.1	Fundamentals of Drug Discovery and CADD Frameworks	5

- Drug discovery lifecycle: target identification to preclinical trials
- Role of CADD in reducing cost and accelerating timelines
- Types of CADD: Structure-based vs Ligand-based
- Overview of small molecules, biologics, and chemical libraries

MODULE NO.2	Structure-Based Drug Design (SBDD)	10

- Protein structure prediction (AlphaFold, homology modeling)
- Molecular docking (AutoDock, Glide) and scoring functions
- Binding site identification and pharmacophore modeling
- Virtual screening and structure optimization

MODULE NO.3	Ligand-Based Drug Design (LBDD) and QSAR Modeling	8

- Ligand similarity search and pharmacophore alignment
- Quantitative Structure–Activity Relationship (QSAR): linear vs nonlinear
- Molecular descriptors and fingerprints (ECFP, MACCS)
- Model validation and performance metrics (R^2, RMSE, ROC-AUC)

MODULE NO.4	Machine Learning and Deep Learning for Drug Discovery	10

- ML algorithms: k-NN, Random Forest, SVM, Gradient Boosting

- Deep learning architectures (MLPs, CNNs, Graph Neural Networks)
- Predictive modeling for ADMET properties and bioactivity
- Autoencoders, GANs, and generative models for molecule design

MODULE NO.5	**AI-Driven CADD Pipelines and Case Studies**	**12**

- Introduction to relational databases
- SELECT, WHERE, GROUP BY, JOIN operations
- Filtering, sorting, aggregations
- Biomedical data querying examples (e.g., patient records, expression tables)
- Using SQL in Python/R pipelines

REFERENCES

1	Patrick, G. L. (2013), *An Introduction to Medicinal Chemistry*, 5th Edition, Oxford University Press, Oxford.
2	Baron, R. (Ed.) (2006), *Computational Drug Discovery and Design*, Springer Protocols, New York.
3	Jhoti, H., & Leach, A. R. (2007), *Structure-Based Drug Discovery*, 1st Edition, Wiley, Chichester.
4	Leach, A. R. (2001), *Molecular Modelling: Principles and Applications*, 2nd Edition, Pearson Education, Essex

2503DDC2T3	Responsible AI, Ethics and Policy in Healthcare		Credits (3)	Hours (45)
COURSE OBJECTIVES: TO ENABLE THE STUDENTS TO-				
1	To introduce foundational ethical principles applicable to the development and deployment of AI in healthcare.			
2	To help learners identify algorithmic bias, safety risks, and propose mitigation strategies in clinical AI applications.			
3	To familiarize students with major regulatory frameworks and compliance requirements for health data and AI products.			
4	To encourage responsible governance practices across the lifecycle of AI systems in healthcare.			
5	To instill integrity, reproducibility, and FAIR principles in AI-driven biomedical research.			
COURSE OUTCOMES: THE STUDENTS WILL BE ABLE TO-				
1	Apply ethical reasoning to real-world challenges in health AI design and deployment			
2	Detects and mitigate sources of bias and safety risks in AI algorithms used in clinical settings.			
3	Navigate legal and regulatory standards such as HIPAA, GDPR, and FDA guidelines.			
4	Implement governance models for AI-based medical products ensuring accountability and transparency.			
5	Conduct responsible biomedical AI research adhering to reproducibility and ethical publication standards.			

COURSE CONTENT

MODULE NO.1	Ethical Foundations of AI in Health Systems	8

- Core principles: fairness, accountability, transparency (FAT)
- Beneficence, non-maleficence, autonomy, and justice in clinical AI
- Real-world ethical dilemmas in algorithmic decision-making
- Case studies: AI in diagnosis, triage, and insurance

MODULE NO.2	Bias, Safety, and Algorithmic Harms in Clinical Settings	8

- Sources of bias: data, modeling, deployment
- Bias detection and mitigation techniques
- Clinical safety risks: false positives, overdiagnosis, automation bias
- Evaluating model impact across subpopulations

MODULE NO.3	Health Data Privacy, Consent, and Security	8

- Data sensitivity and privacy in EHR, imaging, and genomics
- Consent frameworks: opt-in vs opt-out
- Security best practices in health data pipelines
- Case laws and high-profile data breaches

MODULE NO.4	Regulatory Frameworks Governing Health AI	8

- HIPAA (US), GDPR (EU), DPA (India) for health data use

● FDA guidelines for software as a medical device (SaMD) ● CE marking, MHRA, WHO digital health policies ● Risk classification, compliance, and audits		
MODULE NO.5	**Responsible Governance of AI Products**	**8**
● Human-in-the-loop vs AI-autonomous systems ● Transparency in algorithm deployment ● Internal review boards (IRBs), AI ethics boards, and data stewardship ● Lifecycle governance: design, deployment, and decommissioning		
MODULE NO.6	**Scientific Integrity, FAIR Principles & Research Ethics**	**5**
● Reproducibility crisis and causes in AI health research ● FAIR data principles: Findable, Accessible, Interoperable, Reusable ● Systematic reviews and meta-analysis: standards and checklists (PRISMA) ● Research misconduct, authorship, and conflict of interest		

REFERENCES	
1	O'Neil, C. (2016), *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown Publishing Group, New York.
2	Kearns, M., & Roth, A. (2019), *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*, Oxford University Press, Oxford.

2503DDS2T1	Multi-Modality in Healthcare Applications	Credits (3)	Hours (45)
COURSE OBJECTIVES: TO ENABLE THE STUDENTS TO-			
1	To introduce the principles and motivations for using multi-modal data in clinical AI.		
2	To equip learners with knowledge of fusion strategies for combining diverse biomedical modalities.		
3	To develop practical skills in building joint models and shared representations across text, image, omics, and time-series data.		
4	To apply self-supervised and contrastive learning methods for multi-modal representation learning.		
5	To translate theoretical knowledge into real-world use cases in oncology, precision medicine, and critical care.		
COURSE OUTCOMES: THE STUDENTS WILL BE ABLE TO-			
1	Understand and compare multi-modal data fusion techniques and their applications in healthcare.		
2	Design and implement models that jointly process clinical text, medical imaging, genomics, and sensor data.		
3	Apply contrastive and self-supervised learning to improve performance in data-sparse environments.		
4	Learn shared embedding strategies for cross-modal healthcare data integration.		
5	Build AI models that contribute to real-world precision medicine and patient-centered decision systems.		

COURSE CONTENT

MODULE NO.1	Foundations of Multi-Modal Learning in Healthcare	10

- Multi-modal AI and Its role in medicine
- Overview of healthcare modalities: clinical notes, images, omics, vitals
- Challenges of heterogeneity, scale, and missing modalities
- Use cases in integrated diagnostics and decision support

MODULE NO.2	Data Fusion Techniques and Modeling Strategies	5

- Early fusion: concatenation of raw features
- Late fusion: ensemble and decision-level integration
- Hybrid fusion: attention and gating mechanisms
- Best practices in fusion architecture selection

MODULE NO.3	Joint Modeling of Heterogeneous Health Data	8

- Multi-input neural networks and modality-specific encoders
- Joint modeling of text (clinical notes), image (radiology), omics, and time-series (vitals
- Tensor fusion networks, cross-modal transformers
- Handling asynchronous and missing data across modalities

MODULE NO.4	Representation Learning for Multi-Modal Embeddings	8

- Autoencoders and shared latent spaces
- Modality-specific encoders and shared projection heads

- Alignment and co-embedding strategies
- Zero-shot and few-shot learning across unseen combinations

MODULE NO.5	Contrastive and Self-Supervised Learning Techniques	8

- SimCLR, MoCo, BYOL for modality representation
- Contrastive learning for image–text and omics–clinical pairs
- Pretraining with unlabeled medical data
- Self-supervised learning in data-scarce clinical settings

MODULE NO.6	Applications in Precision Health and Critical Care	6

- Oncology: integrating omics + imaging for treatment stratification
- Critical care: vitals + EHR + bedside imagery
- Precision medicine: population health + genomics + lifestyle + monitoring
- Multi-modal digital twin for patient simulation

REFERENCES

1	Zhang, S., et al. (2022), *Multimodal Representation Learning*, Springer, Cham.
2	Barh, D. (Ed.) (2020), *Artificial Intelligence in Precision Health*, Academic Press, Cambridge.
3	Author(s). (2015), *Data Integration in Biological Research: An Overview*, *Bioinformatics*, 31(8), 1301-1309. PMCID: PMC4557916, PMID: 26336651.
4	Kwoji, I. D., Aiyegoro, O. A., Okpeku, M., et al. (2023), *'Multi-omics' Data Integration: Applications in Probiotics Studies*, *npj Science of Food*, 7, 25. DOI: 10.1038/s41538-023-00199-x
5	**Stanojevic, S., et al. (2022),** *Computational Methods for Single-Cell Multi-Omics Integration and Alignment*, *arXiv preprint*, arXiv:2201.06725.
6	**Jiang, Y., et al. (2023),** *A review of multi-omics data integration through deep learning*, *Frontiers in Genetics*, 14, 1199087. DOI: 10.3389/fgene.2023.1199087

2503DDS2T2	Biomedical Time Series and Sensor Data	Credits (3)	Hours (45)
COURSE OBJECTIVES: TO ENABLE THE STUDENTS TO-			
1	To introduce biomedical signal types and the fundamentals of sensor-based data acquisition.		
2	To teach preprocessing and feature extraction methods for robust analysis of biosignals.		
3	To equip students with classical and deep learning-based time series modeling techniques.		
4	To analyze longitudinal and repeated health measurements using statistical and ML methods.		
5	To develop skills for forecasting and detecting critical events from temporal sensor data.		
COURSE OUTCOMES: THE STUDENTS WILL BE ABLE TO-			
1	Understand and process various biomedical signals captured from sensors and wearable devices.		
2	Extract meaningful features from biosignals using signal processing techniques.		
3	Implement ARIMA and deep learning models (RNN, LSTM) to model health-related time series data.		
4	Analyze longitudinal health records for personalized health monitoring.		
5	Forecast patient vitals and detect anomalies in real-time sensor streams for clinical decision support.		

COURSE CONTENT

MODULE NO.1	Foundations of Biomedical Signal Acquisition and Interpretation	8

- Overview of biomedical sensors: EEG, ECG, EMG, PPG, blood pressure, respiration
- Characteristics of biosignals: frequency, amplitude, noise
- Sampling theory and Nyquist-Shannon theorem
- Introduction to wearable sensors and IoT in health

MODULE NO.2	Signal Preprocessing and Feature Engineering	8

- Denoising techniques: moving average, Savitzky-Golay, wavelet transforms
- Baseline correction and signal smoothing
- Time-domain and frequency-domain feature extraction
- Feature selection for physiological time series

MODULE NO.3	Classical Time Series Modeling in Healthcare	8

- AR, MA, ARMA, ARIMA models
- Stationarity, differencing, and seasonal decomposition
- Time series cross-validation
- Use cases: heart rate trend prediction, glucose level modeling

MODULE NO.4	Deep Learning for Temporal Biomedical Data	8

- RNN, GRU, LSTM architectures for sequential data
- Attention mechanisms and temporal convolution
- Multivariate time series modeling
- Applications in disease progression and ICU prediction

MODULE NO.5	Longitudinal and Multi-visit Health Data Analysis	8

- Introduction to longitudinal study designs
- Modeling repeated measures and intra-subject correlation
- Mixed-effect models vs deep learning for patient trajectories
- Use cases: vitals tracking, mental health monitoring

MODULE NO.6	Forecasting and Anomaly Detection in Sensor Data	5

- Forecasting metrics (MAPE, RMSE, MAE)
- Multistep forecasting strategies
- Unsupervised anomaly detection (autoencoders, isolation forest)
- Real-time monitoring applications (fall detection, cardiac anomalies)

REFERENCES

1	Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011), *Applied Longitudinal Analysis*, 2nd Edition, Wiley, New Jersey.
2	Brockwell, P. J., & Davis, R. A. (2016), *Introduction to Time Series and Forecasting*, 3rd Edition, Springer, New York.
3	*ECG-based machine-learning algorithms for heartbeat classification*, *Nature Scientific Reports*, 2021. DOI: 10.1038/s41598-021-97118-5
4	Zhang, G. P. (2003), *Time series forecasting using a hybrid ARIMA and neural network model*, *Neurocomputing*, 50, 159–175. DOI: 10.1016/S0925-2312(01)00702-5
5	Esteban, C., et al. (2016), *Predicting clinical events by combining static and dynamic information using recurrent neural networks*, *Proceedings of the IEEE International Conference on Healthcare Informatics (ICHI)*, 2016, 183-190. DOI: 10.1109/ICHI.2016.34
6	Shashikumar, S. P., et al. (2017), *Early sepsis detection in ICU patients using multinomial LSTM networks*, *arXiv preprint*, arXiv:1705.10516.

2503DDK2P1	**Project 1 OR Project 2**	**Credits (3)**	**Hours**

2503DDK2P2	**Capstone Project**	**Credits (6)**	**Hours**