MAHATMA GANDHI UNIVERSITY

Kottayam, Kerala

Undergraduate Programmes (HONOURS)
2025 Admission Onwards

SYLLABUS			
SIGNATURE COURSE			
Name of the College	St. Thomas College, Ranni		
Faculty/ Discipline	Chemistry		
Programme	BSc (Hons) Chemistry		
Course Coordinator	Haripriya V		
Contributors	Bincy Francis, Chinju Jacob, Swetha C		
Course Name	Data handling in AI & Cheminformatics		
Type of Course	DSE		
Specialization title	Artificial Intelligence and Cheminformatics		
Course Code	MG3DSECHEA04		
Course Level	200		
Course Summary	This course provides a foundational understanding of how data is generated, represented, processed, and utilized in artificial intelligence applications within the field of chemistry. It covers the fundamentals of cheminformatics, including its evolution, scope, and applications in chemical sciences.		
Semester	3	**Credits**	4
Course Details			
Pre-requisites, if any			

Learning Approach	Lecture	Total Hours
Credits	4	60

Course Outcomes (CO)

Number of COs		**4**	
CO No.	**Expected Course Outcome**	**Learning Domains ***	**PO No**
1	Explain the fundamental concepts of cheminformatics, its evolution, and its role in modern scientific research and AI applications.	U	PO1, PO2, PO10
2	Classify and interpret different types, structures, formats, and sources of data, including chemical databases relevant to AI and cheminformatics.	A	PO2, PO3
3	Represent chemical information using standard molecular representations and interpret to identify patterns or errors.	A	PO2, PO4, PO10
4	Apply key steps in the AI data pipeline, including data collection, annotation, cleaning, preprocessing, and transformation, to prepare datasets for machine learning models.	E	PO1, PO2, PO3

**Remember (K), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)*

CO-PO Articulation Matrix

CO/PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10
CO 1	3	3	-	-	-	-	-	-	-	3
CO 2	-	3	3	-	-	-	-	-	-	-
CO 3	-	3	-	3	-	-	-	-	-	2
CO 4	3	3	3	-	-	-	-	-	-	-

'0' is No Correlation, '1' is Slight Correlation (Low level), '2' is Moderate Correlation (Medium level) and '3' is Substantial Correlation (High level).

Course Content

Content for Classroom transaction (Units)

Module	Units	Course Description	Hrs	CO No.
1	Introduction to cheminformatics			
	1.1	Paradigm of science-first, second, third, fourth.	2	["1"]
	1.2	Introduction to Cheminformatics:-History and evolution of cheminformatics, Need and significance of cheminformatics, scope of cheminformatics-drug designing, clinical research, synthetic chemistry and nanotechnology, Chemo vs bioinformatics, Tools for cheminformatics.	3	["1"]
2	Foundations of Data – data Types, data bases , Ethics and Utility in Building Applications using AI			
	2.1	Importance of data in building AI applications: Data as the fuel for AI, Role of big data in training AI models. Conceptual Foundations of Data: Data vs. Information vs. Knowledge. Structure of Data: Structured, Semi-Structured, and Unstructured Data.	4	["2"]
	2.2	Modalities of Data: Text, Image, Audio, Video, Tabular, Time-Series, and Spatial Data. Formats of Data: Text Formats (CSV, JSON, XML), Image Formats (JPEG, GIF, PNG),Audio/Video (MP3, WAV, MP4, AVI).	4	["2"]
	2.3	Data types-Numerical ,Categorical, Nominal, Ordinal, Date and time, Boolean data type, formats of data.	3	["2"]
	2.4	Databases-Concept, types of databases-flat file data base, relational database, primary vs secondary databases Public chemical data base-PubChem, chem spider, ChEBI, ChEMBL , Drug bank, Protein data Bank[elementary idea only].	4	["2"]
3	Molecular representations of chemical Data			
	3.1	Chemical representation for cheminformatics-Line notation- WLN,ROSDAL,SLN (SMILES,INCHI,SMARTS,SMIRKS).	5	["3"]
	3.2	Matrix representations- (Adjacency, Distance, incidence, Bond,) Representation of 3D structure- Z matrix, Connection table, Visualization of adjacency and distance matrices of benzene and ethanol.	5	["3"]
	3.3	various file formats- Structure of mol file and sdf file, comparison, Conversion of data format-Open Babel.	5	["3"]
	3.4	Representation of molecules in SMILES and INCHI benzene, ethanol, cyclohexane, formaldehyde, butane, Connection table for cis –trans butene and benzene.	5	["3"]

Module	Units	Course Description	Hrs	CO No.
4	The AI Data Pipeline: From Collection to Model Readiness			
	4.1	The AI Data Pipeline: Stages and Components: Key Stages (Data Collection, Annotation, Preprocessing, Splitting, Feeding into AI Models).	2	["4"]
	4.2	Data Collection Methods for AI: Manual Input (Surveys, forms, human-curated entries), Sensors & IoT Devices (Real-time data from physical environments), System Logs & Transactions, Web Scraping (Automated extraction from websites), APIs (Structured data access from external platforms).	3	["4"]
	4.3	Data Annotation and Labelling: Manual vs Automated Annotation; Types of Annotation: Classification, Bounding Boxes, Segmentation, Transcription, Named Entity Recognition (NER).	5	["4"]
	4.4	Data Cleaning and Preprocessing: Importance of data cleaning; Understanding “Dirty” Data: Missing Values, Duplicates, Incorrect Formats, Outliers, Noise; Steps in Data Cleaning: Identify Issues, Handle Errors (Imputation, Removal), Validate Cleaned Data Data Splitting: Splitting data into training set and test set.	5	["4"]
	4.5	Data Transformation Techniques: Normalization, Transformation, Feature scaling – Min-Max, Standardization (Z-score), Robust scaling, Feature Engineering(Conceptual).	5	["4"]

Teaching and Learning Approach	**Classroom Procedure (Mode of transaction)** Blended and student-centric approach (lectures, Power Point presentations, quizzes) ,flipped class room, interactive learning through discussions, peer teaching, demonstrations and hands on training, problem solving sessions.

Assessment Types	**MODE OF ASSESSMENT** Mode of Assessment: Theory **A. Continuous Comprehensive Assessment (CCA)** **• Theory – 30 Marks** Quiz, Assignment, Problem Based Test, Written Exam **B. End Semester Evaluation (ESE)** **• Theory – 70 Marks** Assessment Methods – Written Examination Duration of Examination – 2.00 Hrs Pattern of examination for Theory – Non-MCQ Different parts of written examination – Part - A , B , C Answer Type: ◦ PART - A ◦ MCQ - (25 out of 25) – 25 × 1 = 25 ◦ PART - B ◦ Short answer - (5 out of 8) – 5 × 3 = 15 ◦ PART - C ◦ Essays - (2 out of 4) – 2 × 15 = 30

References

- 1. Cheminformatics and Bioinformatics at the Interface with Systems Biology – Aman Chandra Kaushik, Aamir Mehmood, Dongqing Wei, Springer, 2021. 2. Artificial Intelligence for Chemical Sciences: Concepts, Models and Applications – Shrikaant Kulkarni, Shashikant Bhandari, Dushyant Varshney, P. William, Elsevier, 2023. 3. Chemoinformatics: A Textbook – Johann Gasteiger and Thomas Engel, Wiley-VCH, 2003. 4. An Introduction to Chemoinformatics – Andrew R. Leach and Valerie J. Gillet, Springer, 2007. 5. Exploring Chemistry with Electronic Structure Methods – James B. Foresman and A. Eleen Frisch, Gaussian Inc., 2015 (3rd Edition). 6. Handbook of Chemoinformatics – Johann Gasteiger, Wiley-VCH, 2003. 7. Data Mining:

Concepts and Techniques – Jiawei Han, Micheline Kamber, Jian Pei, Morgan Kaufmann (Elsevier), 2011 (3rd Edition).

Suggested Readings

- 1. Managing and Mining Graph Data – Charu C. Aggarwal and Haixun Wang, Springer, 2010. 2. Chemoinformatics: Concepts, Methods, and Tools for Drug Discovery – Johann Gasteiger and Thomas Engel, Wiley-VCH, 2003. 3. Chemical Graph Theory – Nenad Trinajstić, CRC Press, 1992 (2nd Edition). 4. Introduction to Chemoinformatics – Roberto Todeschini and Viviana Consonni, Wiley-VCH, 2009.

Affidavit

- We, St. Thomas College, Ranni and Haripriya V, retain the copyright of this syllabus and expressly prohibit its distribution in complete form to any institution outside our own.
- We, St. Thomas College, Ranni, agree to appoint a new course coordinator for the proposed Artificial Intelligence and Cheminformatics in the event of the unavailability of the currently nominated coordinator. This appointment will ensure the continued coordination of course delivery, assessments, and all related academic responsibilities necessary for the successful implementation of the specialization, for as long as the college offers this programme.
- We, St. Thomas College, Ranni and Haripriya V, declare that no part of this signature course submitted here for approval has been taken from the course content developed by, or from any of the course titles prepared by, the BoS/expert committee in the same discipline under our University.

MAHATMA GANDHI UNIVERSITY

Kottayam, Kerala

Undergraduate Programmes (HONOURS)
2025 Admission Onwards

SYLLABUS			
SIGNATURE COURSE			
Name of the College	St. Thomas College, Ranni		
Faculty/ Discipline	Chemistry		
Programme	BSc (Hons) Chemistry		
Course Coordinator	Haripriya V		
Contributors	Bincy Francis, Chinju Jacob, Swetha C		
Course Name	Cheminformatics: From Molecular Simulations to Artificial Intelligence		
Type of Course	DSE		
Specialization title	Artificial Intelligence and Cheminformatics		
Course Code	MG4DSECHEA04		
Course Level	200		
Course Summary	This course dives into computational foundations and cutting-edge AI for cheminformatics, blending simulations, math, quantum ML, and LLMs. It builds skills for molecular analysis, data handling, and innovative chemical modeling, tailored for research in drug discovery and materials.		
Semester	4	**Credits**	4
Course Details	**Learning Approach** / **Lecture** / **Total Hours** Credits / 4 / 60		
Pre-requisites, if any			

Course Outcomes (CO)

Number of COs		**4**	
CO No.	**Expected Course Outcome**	**Learning Domains ***	**PO No**
1	Simulate and analyze MD trajectories for conformational/ interaction studies.	U	PO1, PO2, PO3, PO10
2	Apply set theory, combinatorics, and graph algorithms to chemical datasets and library design.	U	PO1, PO2, PO3
3	Introduce the theory and practice of neural networks and show how they are applied to chemical data K	K	PO1, PO2, PO3
4	Implement quantum ML and LLM for feature extraction/property prediction and compare with classical methods.	U	PO1, PO2, PO3, PO9, PO10

**Remember (K), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)*

CO-PO Articulation Matrix

CO/PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10
CO 1	3	3	3	-	-	-	-	-	-	3
CO 2	3	3	3	-	-	-	-	-	-	-
CO 3	3	3	3	-	-	-	-	-	-	-
CO 4	3	3	3	-	-	-	-	-	3	3

'0' is No Correlation, '1' is Slight Correlation (Low level), '2' is Moderate Correlation (Medium level) and '3' is Substantial Correlation (High level).

Course Content

Content for Classroom transaction (Units)

Module	Units	Course Description	Hrs	CO No.
1	Discrete Mathematics for Cheminformatics			
	1.1	Set Theory in Chemical Data Analysis: Sets and subsets for molecular classification, operations (union, intersection, difference) for filtering chemical datasets, applications in database querying.	5	["1"]
	1.2	Combinatorial Chemistry: Permutations and combinations for molecular library design, counting molecular configurations, applications in virtual screening.	5	["1"]
	1.3	Advanced Graph Theory: Graph traversal algorithms (DFS, BFS) for molecular path analysis, Application in similarity search.	5	["1"]
2	Molecular Dynamics in Cheminformatics			
	2.1	Introduction to Molecular Dynamics: Principles of molecular dynamics simulations in cheminformatics, Force fields for MD simulations, Applications in studying molecular interactions ,Geometry optimisation and conformational analysis. GUI for MD silmulations- webmo, Avogadro .	5	["2"]
	2.2	Molecular Dynamics Workflow: Preparing molecular systems for MD simulations, Running and analysing MD trajectories using Python tools (e.g., MD Analysis), Visualizing molecular dynamics results for cheminformatics, applications. Hands-On Training: Prepare a small molecule (e.g., benzene) for MD simulation using Python, Run a short MD simulation and extract trajectory data (e.g., RMSD, radius of gyration).	5	["2"]
	2.3	Critical Thinking and Introduction to Logic: Definition and importance of critical thinking, arguments and fallacies, propositional logic (statements, connectives, truth tables), predicate logic (quantifiers: ∀, ∃).	5	["2"]

Module	Units	Course Description	Hrs	CO No.
3	Neural networks in cheminformatics			
	3.1	Neural network basics-Biological inspiration and artificial neurons, Perceptron model, Layers, weights, bias, and activation functions, Forward propagation and output interpretation, artificial neural networks.	3	["3"]
	3.2	Convolutional neural networks (CNN): Using molecular images,1D and 2D convolution concepts, Advantages and limitations for chemical tasks, autoencoders.	2	["3"]
	3.3	Recurrent neural networks (RNN) :Sequential modeling of SMILES, RNN, LSTM, and RU architectures, Sequence generation and prediction.	2	["3"]
	3.4	Graph neural network(GNN): Representing molecules as graphs for ML, GNN architectures for chemical data analysis, Use cases in molecular similarity and substructure prediction. Knowledge graph., Self organising maps .	3	["3"]
	3.5	5Chemical problems solved by neural networks-Property prediction, Bioactivity classification, Molecular similarity, De novo molecule design, Reaction prediction, ADMET modelling.	5	["3"]
4	Emerging approaches in cheminformatics			
	4.1	Introduction to Quantum Machine Learning: Principles of quantum computing for ML applications, Quantum advantages in cheminformatics (e.g., enhanced feature spaces), Overview of quantum algorithms for chemical data.	3	["4"]
	4.2	Quantum Algorithms for Molecular Property Prediction: Variational Quantum Eigensolver (VQE) for molecular energy estimation, Quantum kernel methods for property prediction, Applications in small molecule analysis.	5	["4"]
	4.3	. Quantum-Enhanced Feature Extraction: Quantum feature maps for chemical datasets, Hybrid quantum-classical models for classification, Comparing classical and quantum feature extraction.	5	["4"]
	4.4	Quantum Neural Networks: Architecture of quantum neural networks (QNNs), Training QNNs for cheminformatics tasks, Applications in molecular similarity and clustering feasibility.	2	["4"]

Teaching and Learning Approach	**Classroom Procedure (Mode of transaction)** Lecture sessions (Interactive board, Power Point presentations) ,flipped class room, interactive learning through discussions, peer teaching, Invited lectures, demonstrations.

Assessment Types

MODE OF ASSESSMENT
Mode of Assessment: Theory

A. Continuous Comprehensive Assessment (CCA)
• Theory – 30 Marks
Assignment, MCQ, Written Exam

B. End Semester Evaluation (ESE)
• Theory – 70 Marks
Assessment Methods – Written Examination
Duration of Examination – 2.00 Hrs
Pattern of examination for Theory – Non-MCQ
Different parts of written examination – Part - A , B , C
Answer Type:
- PART - A
- MCQ - (25 out of 25) – 25 × 1 = 25
- PART - B
- Short answer - (5 out of 8) – 5 × 3 = 15
- PART - C
- Essays - (2 out of 4) – 2 × 15 = 30

References

- 1. Cheminformatics and Its Applications – Amalia Stefaniu, CRC Press, 2016. 2. Computational Chemistry and Molecular Modelling: Principles and Applications – K. I. Ramachandran, G. Deepa, K. Namboori, Springer, 2008. 3. Discrete Mathematics and Its Applications – Kenneth H. Rosen, McGraw-Hill Education, 2012 (7th Edition). 4. Graph Theory – Reinhard Diestel, Springer, 2017 (5th Edition). 5. Understanding Molecular Simulation – Daan Frenkel and Berend Smit, Academic Press (Elsevier), 2002 (2nd Edition). 6. Molecular Modelling: Principles and Applications – Andrew R. Leach, Pearson, 2001 (2nd Edition). 7. Neural Networks and Deep Learning – Charu C. Aggarwal, Springer, 2018. 8. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow – Aurélien Géron, O'Reilly Media, 2019 (2nd Edition). 9. Chemoinformatics: Concepts, Methods, and Tools for Drug Discovery – Johann Gasteiger and Thomas Engel, Wiley-VCH, 2003. 10. Quantum Machine Learning – Peter Wittek, Academic Press, 2014; Supervised Learning with Quantum Computers – Maria Schuld and Francesco Petruccione, Springer, 2018. 11. Quantum Computation and Quantum Information – Michael A. Nielsen and Isaac L. Chuang, Cambridge University Press, 2010 (10th Anniversary Edition).

Suggested Readings

- 1. An Introduction to Chemoinformatics – Andrew R. Leach and Valerie J. Gillet, Springer, 2007 (Revised Edition). 2. Pattern Recognition and Machine Learning – Christopher M. Bishop, Springer, 2006. 3. Data-Driven Science and Engineering – Steven L. Brunton and J. Nathan Kutz, Cambridge University Press, 2019. 4. Elements of Discrete Mathematics – C. L. Liu, McGraw-Hill, 1985. 5. Deep Learning – Ian Goodfellow, Yoshua Bengio, and Aaron Courville, MIT Press, 2016.

Affidavit

- We, St. Thomas College, Ranni and Haripriya V, retain the copyright of this syllabus and expressly prohibit its distribution in complete form to any institution outside our own.
- We, St. Thomas College, Ranni, agree to appoint a new course coordinator for the proposed Artificial Intelligence and Cheminformatics in the event of the unavailability of the currently nominated coordinator. This appointment will ensure the continued coordination of course delivery, assessments, and all related academic responsibilities necessary for the successful implementation of the specialization, for as long as the college offers this programme.
- We, St. Thomas College, Ranni and Haripriya V, declare that no part of this signature course submitted here for approval has been taken from the course content developed by, or from any of the course titles prepared by, the BoS/expert committee in the same discipline under our University.

MAHATMA GANDHI UNIVERSITY

Kottayam, Kerala

Undergraduate Programmes (HONOURS)
2025 Admission Onwards

SYLLABUS			
SIGNATURE COURSE			
Name of the College	St. Thomas College, Ranni		
Faculty/ Discipline	Chemistry		
Programme	BSc (Hons) Chemistry		
Course Coordinator	Haripriya V		
Contributors	Bincy Francis, Chinju Jacob, Swetha C		
Course Name	Fundamentals of AI and Machine learning		
Type of Course	DSE		
Specialization title	Artificial Intelligence and Cheminformatics		
Course Code	MG5DSECHEA04		
Course Level	300		
Course Summary	This course provides a comprehensive introduction to artificial intelligence and machine learning, emphasizing their fundamental concepts, techniques, and applications in the chemical sciences. It covers key methodologies such as supervised and unsupervised learning, along with tools and platforms including KNIME, Orange, and Google Colab for developing data-driven models. The course also focuses on cheminformatics concepts such as molecular descriptors, fingerprints, chemical space analysis, and similarity searching for effective chemical data analysis and prediction.		
Semester	5	**Credits**	4
Course Details	**Learning Approach** / **Lecture** / **Total Hours** Credits / 4 / 60		
Pre-requisites, if any			

Course Outcomes (CO)

Number of COs		**4**	
CO No.	**Expected Course Outcome**	**Learning Domains ***	**PO No**
1	Explain fundamental concepts, types, and key technologies of artificial intelligence and their applications in chemical sciences.	U	PO1, PO2, PO10
2	Apply machine learning techniques, including decision trees and support vector machines, for solving chemical data problems.	A	PO2, PO3, PO5
3	Utilize AI platforms and tools (e.g., KNIME, Orange, Google Colab) to build and implement data-driven applications.	A	PO2, PO4, PO5, PO9, PO10
4	Analyze chemical data using molecular descriptors, fingerprints, and similarity measures for property prediction and chemical space exploration.	E	PO1, PO2, PO3, PO10

**Remember (K), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)*

CO-PO Articulation Matrix

CO/PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10
CO 1	3	3	-	-	-	-	-	-	-	3
CO 2	-	3	3	-	3	-	-	-	-	-
CO 3	-	3	-	3	3	-	-	-	3	3
CO 4	3	3	3	-	-	-	-	-	-	3

'0' is No Correlation, '1' is Slight Correlation (Low level), '2' is Moderate Correlation (Medium level) and '3' is Substantial Correlation (High level).

Course Content

Content for Classroom transaction (Units)

Module	Units	Course Description	Hrs	CO No.
1	Introduction to Artificial Intelligence			
	1.1	Introduction to AI ,ML and deep learning.	2	["1"]
	1.2	Types (narrow, general, super intelligent) and components of AI(data, algorithm, model, computing power).	2	["1"]
	1.3	Key AI technologies-Natural language processing (NLP), Neural networks, computer vision, generative AI, expert systems, reinforcement learning Applications: Molecular design & property prediction, Reaction optimization & synthesis, Analytical chemistry & data analysis.	6	["1"]
2	Basics of Machine Learning			
	2.1	Machine learning: Definition and types (supervised, unsupervised, reinforcement), ML workflow: data preprocessing, feature selection, model building (training, evaluation, screening), applications-QSAR modelling, ADMET prediction, virtual screening.	5	["2"]
	2.2	Decision Trees: Definition, tree structure (root, nodes, leaves), regression leaf prediction rule, splitting criteria (Gini index, entropy, information gain), recursive partitioning. CART (Classification and Regression Trees) vs ID3 , Advantages and disadvantages, Applications in predicting molecular properties, Random forests and gradient boosting (e.g., XG Boost).	5	["2"]
	2.3	Support Vector Machines (SVM): Linear and non-linear SVM models, Hard and soft margins, Kernel functions (e.g., RBF, polynomial) for chemical data, Use in molecular classification and regression.	5	["2"]
	2.4	Model Evaluation and Validation: Performance metrics: accuracy, precision, recall, F1-score, Cross-validation techniques (e.g., k-fold), Addressing overfitting and underfitting in chemical models. Validation: R^2, ROC-AUC,RMSE, cross-validation, Y-randomization.	5	["2"]
3	Infrastructure and Platforms for Building Applications using AI			
	3.1	Hardware used in building AI applications Processors - CPU, GPU, TPU, NPU, Memory - RAM, VRAM, Storage - HDD, SSD Platforms for building applications using AI: Online platforms (Example - Google Auto ML, H2O.ai, Teachable Machine or similar platforms –Elementary idea only) Desktop (No-code/Low code) platforms (Orange, Data Mining, KNIME, Weka, RapidMiner or similar tools-Elementary idea only).Hands on training with KNIME and Orange.	10	["3"]
	3.2	Google Colab for Cheminformatics: Setting up Python environments in Google colab, Running cheminformatics scripts and deploying ML models in the cloud. Advantages , applications.	5	["3"]

Module	Units	Course Description	Hrs	CO No.
4	Chemical space molecular descriptors and fingerprints			
	4.1	Molecular Descriptors and Fingerprints: Types of descriptors: 2D, 3D, constitutional, topological, geometrical, electronic, Generating descriptors using computational tools, Role of descriptors in ML model development, types of molecular fingerprint (structural, path based, circular, pharmacophore based), MACCS keys and Morgan fingerprint, , Hands on training on fingerprint generation with RDKit .	7	["4"]
	4.2	Chemical Space Analysis: Definition and importance of chemical space, Principal Component Analysis (PCA) for dimensionality reduction, introduction to t-SNE, UMAP methods Visualizing chemical space for compound analysis.	3	["4"]
	4.3	Introduction to similarity searching, Similar structure similar property principle, Similarity coefficients (e.g., Tanimoto, Dice),structure –property relationship, Methods for similarity-based compound identification and property prediction.	5	["4"]

Teaching and Learning Approach	**Classroom Procedure (Mode of transaction)** Lecture sessions, Interactive sessions including discussion, demonstrations, peer teaching, ICT enabled learning for conceptual clarity, hand on training to enhance understanding, case studies.

Assessment Types	**MODE OF ASSESSMENT** Mode of Assessment: Theory **A. Continuous Comprehensive Assessment (CCA)** **• Theory – 30 Marks** Quiz, Assignment, MCQ, Written Exam **B. End Semester Evaluation (ESE)** **• Theory – 70 Marks** Assessment Methods – Written Examination Duration of Examination – 2.00 Hrs Pattern of examination for Theory – Non-MCQ Different parts of written examination – Part - A , B , C Answer Type: ◦ PART - A ◦ MCQ - (25 out of 25) – 25 × 1 = 25 ◦ PART - B ◦ Short answer - (5 out of 8) – 5 × 3 = 15 ◦ PART - C ◦ Essays - (2 out of 4) – 2 × 15 = 30

References

- 1. Artificial Intelligence: A Modern Approach – Stuart Russell and Peter Norvig, Pearson, 2020 (4th Edition). 2. Artificial Intelligence – Charu C. Aggarwal, Springer, 2021. 3. Pattern Recognition and Machine Learning – Christopher M. Bishop, Springer, 2006. 4. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow – Aurélien Géron, O'Reilly Media, 2019 (2nd Edition). 5. The Master Algorithm – Pedro Domingos, Basic Books, 2015. 6. Machine Learning Engineering – Andriy Burkov, True Positive Inc., 2020. 7. Introduction to Chemoinformatics – Roberto Todeschini and Viviana Consonni, Wiley-VCH, 2009. 8. Chemoinformatics: Theory, Practice, and Products – Jürgen Bajorath, Springer, 2004. 9. Molecular Descriptors for Chemoinformatics – Roberto Todeschini and Viviana Consonni, Wiley-VCH, 2009 (2nd Edition).

Suggested Readings

- 1. Machine Learning: A Probabilistic Perspective – Kevin P. Murphy, MIT Press, 2012. 2. Data Mining: Practical Machine Learning Tools and Techniques – Ian H. Witten, Eibe Frank, Mark A. Hall, Morgan Kaufmann (Elsevier), 2016 (4th Edition). 3. Deep Learning – Ian Goodfellow, Yoshua Bengio, Aaron Courville, MIT Press, 2016. 4. Designing Machine Learning Systems – Chip Huyen, O'Reilly Media, 2022. 5. Feature Engineering for Machine Learning and Data Analytics – Dong Guozhu and Liu Huan, CRC Press, 2018.

Affidavit

- We, St. Thomas College, Ranni and Haripriya V, retain the copyright of this syllabus and expressly prohibit its distribution in complete form to any institution outside our own.
- We, St. Thomas College, Ranni, agree to appoint a new course coordinator for the proposed Artificial Intelligence and Cheminformatics in the event of the unavailability of the currently nominated coordinator. This appointment will ensure the continued coordination of course delivery, assessments, and all related academic responsibilities necessary for the successful implementation of the specialization, for as long as the college offers this programme.
- We, St. Thomas College, Ranni and Haripriya V, declare that no part of this signature course submitted here for approval has been taken from the course content developed by, or from any of the course titles prepared by, the BoS/expert committee in the same discipline under our University.

MAHATMA GANDHI UNIVERSITY

Kottayam, Kerala

Undergraduate Programmes (HONOURS)
2025 Admission Onwards

SYLLABUS			
SIGNATURE COURSE			
Name of the College	St. Thomas College, Ranni		
Faculty/ Discipline	Chemistry		
Programme	BSc (Hons) Chemistry		
Course Coordinator	Haripriya V		
Contributors	Bincy Francis, Chinju Jacob, Swetha C		
Course Name	Applications of Artificial Intelligence in Chemistry		
Type of Course	DSE		
Specialization title	Artificial Intelligence and Cheminformatics		
Course Code	MG6DSECHEA04		
Course Level	300		
Course Summary	It bridges theory and practice, using Python ecosystems (RDKit, Deep Chem) for chemical data analysis, emphasizing ethical, scalable AI in chemistry subfields like cheminformatics and bioinformatics.		
Semester	6	**Credits**	4
Course Details			
Pre-requisites, if any			

Learning Approach	Lecture	Practical/Lab-P	Total Hours
Credits	3	1	75

Course Outcomes (CO)

Number of COs		4	
CO No.	**Expected Course Outcome**	**Learning Domains ***	**PO No**
1	Articulate AI ethics principles and assess risks in chemical AI systems.	U	PO6, PO7, PO8, PO10
2	Use Python libraries/databases for material property prediction and visualization.	A	PO1, PO2, PO3, PO10
3	Apply AI pipelines for drug target discovery, structure prediction, binding sites, virtual screening, interactions, and ADMET.	A	PO2, PO3, PO5, PO9, PO10
4	Execute practical: descriptor generation, PCA, rule-based filtering, ML toxicity models, and chemical space visualization.	An, S	PO1, PO2, PO4, PO5, PO10

**Remember (K), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)*

CO-PO Articulation Matrix

CO/PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10
CO 1	-	-	-	-	-	3	3	3	-	3
CO 2	3	3	3	-	-	-	-	-	-	3
CO 3	-	3	3	-	3	-	-	-	3	3
CO 4	3	3	-	3	3	-	-	-	-	3

'0' is No Correlation, '1' is Slight Correlation (Low level), '2' is Moderate Correlation (Medium level) and '3' is Substantial Correlation (High level).

Course Content

Content for Classroom transaction (Units)

Module	Units	Course Description	Hrs	CO No.
1	Ethical AI and future trends			
	1.1	1Ethical principles-Introduction to AI ethics, ethical challenges in AI, Ethical data handling and documentation.	3	["1"]
	1.2	Risk of AI deployment-misinformation, hallucination, intellectual property concerns, mitigation strategies.	4	["1"]
	1.3	Ethical risk assessment-Risk matrices for AI systems, National and international AI ethics policies.	3	["1"]
2	AI in Material Science			
	2.1	Introduction to python libraries for cheminformatics- Data handling-pandas, NumPy, property prediction- Scikit learn, RDkit, DeepChem, Descriptor generator-PADEL, Mordred [elementary idea only]	5	["2"]
	2.2	Databases: Scope, contents, data sources and applications of COD (Crystallography Open Database) and ICSD (Inorganic Crystal Structure Database), materials project.	2	["2"]
	2.3	Graphical tools for data visualization in material research: Heat maps, parallel coordination, Radial visualization, glyph plots, Image-Based material Characterization- Open CV, Tensor Flow/ Keras for image classification and segmentation.	3	["2"]
	2.4	Applied AI: Raman & IR spectroscopy, synchrotron facilities, scanning probe microscopy, EM, material design and property prediction.	5	["2"]

Module	Units	Course Description	Hrs	CO No.
3	AI in Drug Designing			
	3.1	Drug target discovery: Limitations of single omics approaches, Pipeline of Multi-omics data integration analyses, Multi-omics databases (Cancer ,Genome Atlas, Proteomics DB,), AI driven bigdata handling in omics (Scalability, data integration, automated feature selection).	3	["3"]
	3.2	Protein Structure prediction: Importance in drug discovery and structural Biology, Experimental methods (X Ray crystallography, NMR, Cryo-EM) and computational methods (Homology modelling, Ab inito modelling, Threading, AI based prediction – Deep learning).	3	["3"]
	3.3	Binding site prediction – Challenges in traditional binding site prediction, Comparitive study of AI based binding site prediction method – Deepsite, P2Rank, Deep BindBC.	2	["3"]
	3.4	Virtual Screening: Evolution (Early computational approaches, molecular docking, high throughput virtual screening, Artificial Intelligence & Machine Learning), AI in virtual screening – Ligand Based Virtual Screening, Structure Based Virtual Screening.	3	["3"]
	3.5	Molecular docking:-Uses ,steps, scoring function, limitations of traditional docking, GNINA-AI enhanced docking, Advantages of AI assisted docking.	2	["3"]
	3.6	Drug-Target interaction Prediction: Feature based models, Similarity based models, GNNs, Deep Learning & Transformer Models.	2	["3"]
	3.7	ADMET prediction: Conventional ADMET prediction methods, AI based ADMET prediction, Common AI models (DruMAP, ADMET AI), AI vs conventional methods.	5	["3"]
4	Practicals			
	4.1	1. Explore materials science datasets (eg. band gap, conductivity) from the materials project database or kaggle using Orange data mining-perform correlation analysis to identify relationships between material properties 2. Perform chemical compound classification using datasets from PubChem or ChEMBL with ChemMine tools 3. Molecular descriptor generation using PADEL and RD kit. 4. Predicting solubility using molecular descriptor (KNIME, Google colab) 5. Performing PCA on a set of molecular descriptors using Python (Scikit-learn). 6. Applying Lipinski's Rule of Five to a set of compounds to predict the drug likeness of a compound. 7. Analyse ADMET properties of FDA approved drug using DruMap-evaluate pharmacokinetic parameters and drug likeness profile 8. Perform AI assisted molecular docking of a selected protein from the RCSB protein Data bank(with a co –crystallized ligand)using google colab-determine binding affinity (docking score) 9. Train a model to predict the toxicity of a compound using Python libraries.	30	["4"]

Teaching and Learning Approach	**Classroom Procedure (Mode of transaction)** Blended and interactive approach combining lectures(Chalk & board, power point presentations),flipped classroom, demonstrations, and group discussions, debates, real-world case studies, project based learning, internships ,Experimental learning (hands-on, skill-oriented) ,problem solving method, collaborative learning. • Toxicity prediction model for small molecules. • Solubility prediction using neural networks. • Drug-likeness classifier. • Multi-task property prediction model.

Assessment Types	**MODE OF ASSESSMENT** Mode of Assessment: Both **A. Continuous Comprehensive Assessment (CCA)** **• Theory – 25 Marks** Quiz, Assignment, MCQ, Written Exam **• Practical – 15 Marks** Lab involvement,Lab report, Written Test **B. End Semester Evaluation (ESE)** **• Theory – 50 Marks** Assessment Methods – Written Examination Duration of Examination – 1.50 Hrs Pattern of examination for Theory – Non-MCQ Different parts of written examination – Part - A , B , C Answer Type: ◦ PART - A ◦ MCQ - (20 out of 20) – 20 × 1 = 20 ◦ PART - B ◦ Short answer - (5 out of 8) – 5 × 3 = 15 ◦ PART - C ◦ Essays - (1 out of 2) – 1 × 15 = 15 **• Practical – 35 Marks** Assessment Methods – practical based assessment(Viva voce-10,Writting Procedure-15,certified Lab report-10) Duration of Examination – 1.00 Hrs

References

- 1. Artificial Intelligence in Drug Discovery – Nathan Brown, Royal Society of Chemistry, 2020. 2. Machine Learning in Chemistry – Edward O. Pyzer-Knapp and Hugh G. Kerber, ACS Publications, 2020. 3. Information Theory and Statistics in Materials Science – Richard LeSar, Cambridge University Press, 2013. 4. Computational Materials Science: An Introduction – June Gunn Lee, CRC Press, 2016. 5. Materials Informatics: Methods, Tools, and Applications – Olexandr Isayev, Wiley, 2022. 6. Data-Driven Materials Science – Turab Lookman, Cambridge University Press, 2016. 7. Drug-Like Properties: Concepts, Structure Design and Methods from ADME to Toxicity Optimization – Edward H. Kerns and Li Di, Academic Press (Elsevier), 2008. 8. Cheminformatics, QSAR and Machine Learning Applications for Novel Drug Development – Kunal Roy, IGI Global, 2020. 9. Artificial Intelligence and Machine Learning in Drug Design and Development – Abhirup Khanna et al., CRC Press, 2021. 10. An Introduction to Medicinal Chemistry – Graham L. Patrick, Oxford University Press, 2017 (5th Edition).

Suggested Readings

- 1. Computational Materials Science: An Introduction – June Gunn Lee, CRC Press, 2016. 2. Molecular Modelling: Principles and Applications – Andrew R. Leach, Pearson, 2001 (2nd Edition). 3. Computer Aided Drug Design – D. B. Singh, Springer, 2019

Affidavit

- We, St. Thomas College, Ranni and Haripriya V, retain the copyright of this syllabus and expressly prohibit its distribution in complete form to any institution outside our own.
- We, St. Thomas College, Ranni, agree to appoint a new course coordinator for the proposed Artificial Intelligence and Cheminformatics in the event of the unavailability of the currently nominated coordinator. This appointment will ensure the continued coordination of course delivery, assessments, and all related academic responsibilities necessary for the successful implementation of the specialization, for as long as the college offers this programme.
- We, St. Thomas College, Ranni and Haripriya V, declare that no part of this signature course submitted here for approval has been taken from the course content developed by, or from any of the course titles prepared by, the BoS/expert committee in the same discipline under our University.