MODULE DESCRIPTION FORM

نموذج وصف المادة الدراسية

| **Module Information**  **معلومات المادة الدراسية** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Module Title** | Pattern Discovery in Bioinformatics | | | | **Module Delivery** | | |
| **Module Type** | Core | | | | * **☒ Theory** * **☐Lecture** * **☒ Lab** * **☐ Tutorial** * **☐ Practical** * **☐ Seminar** | | |
| **Module Code** | BID321 | | | |
| **ECTS Credits** | 6.00 | | | |
| **SWL (hr/sem)** | 150 | | | |
| **Module Level** | | UGx1 3 | **Semester of Delivery** | | | | 6 |
| **Administering Department** | | BID | **College** | BMIC | | | |
| **Module Leader** | Usama Falah | | **e-mail** | osama.falah@uoitc.edu.iq | | | |
| **Module Leader’s Acad. Title** | |  | **Module Leader’s Qualification** | | | |  |
| **Module Tutor** | Name (if available) | | **e-mail** | E-mail | | | |
| **Peer Reviewer Name** | | omar A. M. | **e-mail** | omara.m@uoitc.edu.iq | | | |
| **Scientific Committee Approval Date** | | 18/06/2023 | **Version Number** | | | 1.0 | |

| **Relation with other Modules**  **العلاقة مع المواد الدراسية الأخرى** | | | |
| --- | --- | --- | --- |
| **Prerequisite module** | Image Processing / BMI312,  Bioinformatics Algorithms / BID311 | **Semester** | 5 |
| **Co-requisites module** | None | **Semester** |  |

| **Module Aims, Learning Outcomes and Indicative Contents**  **أهداف المادة الدراسية ونتائج التعلم والمحتويات الإرشادية** | |
| --- | --- |
| **Module Objectives**  **أهداف المادة الدراسية** | The objectives in pattern discovery in bioinformatics are to identify and understand patterns, motifs, and structures in biological data that can provide insights into various biological processes. Here are some common objectives in pattern discovery in bioinformatics:  1. Identification of Sequence Motifs: Discovering short conserved patterns or motifs in DNA, RNA, or protein sequences that are associated with specific biological functions or regulatory elements.  2. Pattern Mining in Gene Expression Data: Identifying patterns and relationships in gene expression data to understand gene regulatory networks, identify co-regulated genes, and uncover biological pathways and processes.  3. Protein Structure Motif Identification: Discovering recurring structural patterns or motifs in protein structures that are important for protein function, folding, and interactions. |
| **Module Learning Outcomes**  **مخرجات التعلم للمادة الدراسية** | The outcomes in pattern discovery in bioinformatics are focused on gaining insights into biological processes, understanding molecular mechanisms, and improving our knowledge of the structure and function of biological molecules. Here are some common outcomes in pattern discovery in bioinformatics:  1. Identification of Novel Patterns and Motifs: The discovery of new patterns and motifs in biological data helps in understanding the underlying mechanisms of biological processes, such as gene regulation, protein-protein interactions, and signaling pathways.  2. Functional Annotation of Genomic Sequences: Identifying patterns in genomic sequences can aid in the annotation of genes and regulatory elements. This information helps in understanding gene function, predicting protein structure and function, and deciphering gene regulatory networks.  3. Characterization of Protein Structures and Functional Sites: Pattern discovery in protein structures can reveal conserved motifs and functional sites, leading to a better understanding of protein folding, molecular interactions, and enzyme catalysis. This information is crucial for drug discovery and protein engineering.  4. Prediction of Regulatory Elements: Discovery of regulatory motifs in DNA sequences can help identify transcription factor binding sites, promoter regions, and other regulatory elements. This knowledge contributes to understanding gene regulation, genetic diseases, and the design of gene therapies.  5. Integration of Diverse Data Types: Pattern discovery allows for the integration of various data types, such as genomics, transcriptomics, proteomics, and epigenomics. Integrating these data sources can provide a comprehensive view of biological systems and reveal novel patterns and associations.  6. Development of Computational Tools and Algorithms: Pattern discovery in bioinformatics often leads to the development of novel computational tools and algorithms. These tools can be used by researchers and analysts to explore and analyze biological data, and they contribute to advancements in the field of bioinformatics.  7. Insights into Biological Mechanisms and Disease Processes: The outcomes of pattern discovery can provide insights into fundamental biological mechanisms, including gene regulation, cellular processes, and disease mechanisms. This knowledge can aid in the development of targeted therapies and personalized medicine.  8. Collaboration and Knowledge Sharing: The outcomes of pattern discovery in bioinformatics often involve collaboration and knowledge sharing among researchers and scientists. Sharing discovered patterns, tools, and methodologies enables the broader scientific community to benefit from the findings and accelerates advancements in the field. |
| **Indicative Contents**  **المحتويات الإرشادية** | Pattern discovery in bioinformatics involves the identification and analysis of meaningful patterns or motifs within biological data, such as DNA or protein sequences. These patterns can provide valuable insights into the structure, function, and evolution of biological molecules. While the specific contents may vary depending on the context and methods employed, here are some indicative contents typically involved in pattern discovery in bioinformatics:  1. Introduction to pattern discovery: Provide an overview of the field of bioinformatics and the importance of pattern discovery in understanding biological processes.  2. Biological sequence data: Discuss different types of biological sequences, such as DNA, RNA, and protein sequences, along with their properties and representations.  3. Motif definition and representation: Define motifs as recurring patterns or sequences of interest within biological data. Explore various approaches for representing motifs, such as consensus sequences, position weight matrices (PWMs), or regular expressions. |

| **Learning and Teaching Strategies**  **استراتيجيات التعلم والتعليم** | |
| --- | --- |
| **Strategies** | In pattern discovery in bioinformatics, several strategies and approaches are employed to identify and analyze meaningful patterns within biological data. Here are some common strategies used in pattern discovery: Statistical approaches: Statistical methods are widely used to identify patterns that are statistically significant. These approaches involve comparing observed patterns with expected patterns based on random chance or background models. Examples include Fisher's exact test, chi-square test, and hypergeometric distribution-based tests.  Sequence alignment: Sequence alignment techniques, such as pairwise alignment or multiple sequence alignment, are used to identify conserved regions or motifs within a set of related sequences. These methods help to identify patterns that may have functional or structural significance. Hidden Markov Models (HMMs): HMMs are probabilistic models that are used to represent patterns or motifs with probabilistic states and transitions. They are particularly useful for finding patterns that have variable lengths or contain gaps. HMM-based algorithms, such as the Baum-Welch algorithm, Viterbi algorithm, and Forward-Backward algorithm, are used for motif discovery. |

| **Student Workload (SWL)**  **الحمل الدراسي للطالب محسوب لـ ١٥ أسبوعا** | | | |
| --- | --- | --- | --- |
| **Structured SWL (h/sem)**  **الحمل الدراسي المنتظم للطالب خلال الفصل** | 64 | **Structured SWL (h/w)**  **الحمل الدراسي المنتظم للطالب أسبوعيا** | 4 |
| **Unstructured SWL (h/sem)**  **الحمل الدراسي غير المنتظم للطالب خلال الفصل** | 86 | **Unstructured SWL (h/w)**  **الحمل الدراسي غير المنتظم للطالب أسبوعيا** | 6 |
| **Total SWL (h/sem)**  **الحمل الدراسي الكلي للطالب خلال الفصل** | **150** | | |

| **Module Evaluation**  **تقييم المادة الدراسية** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **As** | | **Time/Number** | **Weight (Marks)** | **Week Due** | **Relevant Learning Outcome** |
| **Formative assessment** | **Quizzes** | 2 | 10% (10) | 5 and 8 | LO #1, #2 and #8 |
| **Assignments** | 2 | 10% (10) | 2 and 7 | LO #3, #4 and #6, #7 |
| **Projects / Lab.** | 1 | 10% (10) | Continuous | All |
| **Report** | 1 | 10% (10) | 13 | LO #5, #7 and #8 |
| **Summative assessment** | **Midterm Exam** | 2hr | 10% (10) | 7 | LO #1 - #7 |
| **Final Exam** | 3hr | 50% (50) | 16 | All |
| **Total assessment** | | | 100% (100 Marks) |  |  |

| **Delivery Plan (Weekly Syllabus)**  **المنهاج الاسبوعي النظري** | |
| --- | --- |
| **Week** | **Material Covered** |
| **Week 1** | Introduction to Pattern Discovery in Bioinformatics |
| **Week 2** | Sequence Alignment |
| **Week 3** | Hidden Markov Models (HMMs) |
| **Week 4** | Motif Discovery |
| **Week 5** | Regulatory Motifs and Transcription Factor Binding Sites |
| **Week 6** | Gene Expression Analysis |
| **Week 7** | Mid-term Exam |
| **Week 8** | Protein Structure Prediction |
| **Week 9** | Protein-Protein Interaction Networks |
| **Week 10** | Next-Generation Sequencing (NGS) Data Analysis |
| **Week 11** | Machine Learning in Pattern Discovery |
| **Week 12** | Epigenomics and DNA Methylation Analysis |
| **Week 13** | Structural Variant Detection and Analysis |
| **Week 14** | Analysis of SVs in disease studies and population genetics |
| **Week 15** | Review |

| **Delivery Plan (Weekly Lab. Syllabus)**  **المنهاج الاسبوعي للمختبر** | |
| --- | --- |
| **Week** | **Material Covered** |
| **Week 1-2** | Lab 1: Sequence Alignment |
| **Week 3-4** | Lab 2: Hidden Markov Models (HMMs |
| **Week 5-6** | Lab 3: Motif Discovery |
| **Week 7-8** | Lab 4: Gene Expression Analysis |
| **Week 9-10** | Lab 5: Protein-Protein Interaction Networks |
| **Week 11-12** | Lab 6: Structural Variant Detection and Analysis |
| **Week 13-14** | Lab 7: Analysis of SVs in disease studies and population genetics |
| **Week 15** | Review |

| **Learning and Teaching Resources**  **مصادر التعلم والتدريس** | | |
| --- | --- | --- |
|  | **Text** | **Available in the Library?** |
| **Required Texts** | Aggarwal, C. C. (2015). Data Mining: The Textbook. Springer. | No |
| **Recommended Texts** | Eskin, E. (2020). Bioinformatics for Geneticists. CRC Press. | No |

| **Grading Scheme**  **مخطط الدرجات** | | | | |
| --- | --- | --- | --- | --- |
| **Group** | **Grade** | **التقدير** | **Marks %** | **Definition** |
| **Success Group**  **(50 - 100)** | **A -** Excellent | **امتياز** | 90 - 100 | Outstanding Performance |
| **B -** Very Good | **جيد جدا** | 80 - 89 | Above average with some errors |
| **C -** Good | **جيد** | 70 - 79 | Sound work with notable errors |
| **D -** Satisfactory | **متوسط** | 60 - 69 | Fair but with major shortcomings |
| **E -** Sufficient | **مقبول** | 50 - 59 | Work meets minimum criteria |
| **Fail Group**  **(0 – 49)** | **FX –** Fail | **راسب (قيد المعالجة)** | (45-49) | More work required but credit awarded |
| **F –** Fail | **راسب** | (0-44) | Considerable amount of work required |
|  |  |  |  |  |
| **Note:** Marks Decimal places above or below 0.5 will be rounded to the higher or lower full mark (for example a mark of 54.5 will be rounded to 55, whereas a mark of 54.4 will be rounded to 54. The University has a policy NOT to condone "near-pass fails" so the only adjustment to marks awarded by the original marker(s) will be the automatic rounding outlined above. | | | | |