Cancer Prediction Project Plan

By: Group 4
Elaine Liong (ID: 29942357)
Jack Ooi (ID: 29037077)
Vionnie Tan (ID:30092809)

Speaker: Jack

Agenda

1. Literature Review
2. Project Management Plan
3. Project Design
4. Methodology
5. Test Planning

Speaker: Jack

LITERATURE REVIEW

Speaker: Jack

History of Cancer Classification

- Earlier cancer classification techniques (Munir et al, 2019)
  - Menzies method
  - Seven-point detection method
- Newer classification techniques (Munir et al, 2019)
  - Deep neural networks
    - Convolutional Neural Network (CNN)
    - Recurrent Neural Network (RNN)

Speaker: Jack

Issues of Medical Data

- Medical data is hard to obtain (Delen, 2009) & (Kumar et al, 2019)
  - Patient' confidentiality
  - Vast variety (Kumar et al, 2019)
- Investigation of Medical images is hard (Nahid et al, 2018)
  - Complex nature of histopathological imaging

Speaker: Jack

Classification Techniques

- Multiple classification techniques (Gupta et al, 2011)
  - Artificial Neural Network (ANN)
  - Decision Tree (DT)
  - Support Vector Machine (SVM)
- Performance
  - Decision Tree is faster than ANN (Kumar et al, 2014)
  - Different techniques with pros and cons (Gupta et al, 2011)

Speaker: Jack

Image Processing Techniques

- Steps of Image Processing (Ansary et al, 2017) & (Jain et al, 2015)
  - Pre-processing
    - Grayscale extraction
  - Image segmentation
    - Masking operation
  - Feature extraction

Speaker: Jack

Feature extraction techniques

- Multiple deep learning algorithms (Alom et al . 2018)
  - Convolutional Neural Networks (CNN)
  - Alexnet
  - Residual Neural Network (Resnet)
- Performance
  - Resnet -> 84.09 % (Demir et al, n.d)
  - CNN -> 80.8% to 85.6% (Spanhol et al, 2016)
  - Alexnet -> 93.8% to 95.7% (Titoriya, 2019)

Speaker: Jack

PROJECT MANAGEMENT PLAN

Speaker: Elaine

Product Requirements

REQUIREMENTS TRACEABILITY MATRIX					
Project Name:	**Data Mining Technique To Detect Cancer Using Predictive Modelling**				
Project Manager Name:	**Afsaneh Koohestani**				
Project Description:	**Building a predictive model to determine early stages of cancer**				
ID	***Requirements (Functional or Non-Functional)***	***Assumption(s) and/or Customer Need(s)***	***Category***	***Source***	**Status**
001	Image Processing - Find relevant dataset that could be used for testing & training	Source of images are come from a legit source	Functional	Kaggle	In Progress
002	Programming Skills - Understanding topics regarding AI, Machine Learning, Deep Learning, Transfer Learning	Extensive knowledge of these programming skills increases the chance of efficiency of our predictive model	Functional	Online resources such as Stack Overflow, Leetcode, and Monash Units	Planning
003	User Interface - Allow login specific of the health institutions	Our stakeholders have to be able to easily access and understand the interface for them to use the predictive model	Non-functional	Project supervisor	Planning

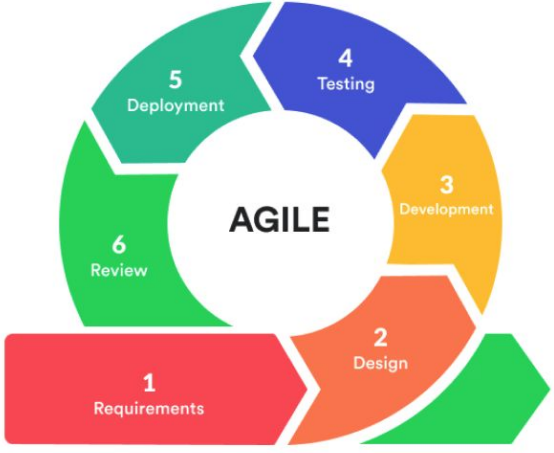
Speaker: Elaine

Product Requirements

004	Stakeholder expectations met	Accuracy of the Model has to be > 98%	Non-functional	Stakeholders	Planning
005	The software system should be integrated with health institutions's API	If we are building on an existing model, then our predictive model must be integrated with the current system.	Functional	Stakeholders	Planning
006	Identify important predictors that have significant impact on successful cancer categorization	Identifying important predictors will allow us to increase model accuracy	Functional	Project supervisor	Planning

Speaker: Elaine

Project Organisation



(InDevLAB, 2020)

Speaker: Elaine

Project Responsibilities

- Building predictive model

- Developing a website

Speaker: Elaine

Risk Management

No.	Rank	Risk	Description	Category	Triggers	Root Cause	Potential Responses	Risk Owner	Probability	Impact	Status	Score
1	6	Losing team members	Team members leaving the team	People risk	A team member decides to leave the team	Team member's personal issue	Consult project manager, redefine task responsibilities for each remaining team member	Team	5%	High	Potential	
2	9	Team members unable to contribute	Team member not able to complete their task responsibilities	People risk	A team member encounters some issues that affects their work	Team member's personal issue	Consult project manager	Team	10%	Medium	Potential	
3	8	Slow decision making / Project Conflicts	Indecisive and not prioritising the success of the project rather for personal gains. Unclear of project objectives and requirements.	Management risk	Clash within personal interests of team members and unsuccessful understanding of given tasks.	Lack of open-mindedness and clarity within team members	Conduct internal meetings with team members and settle on a middle ground.	Team	15%	Medium	Potential	

Speaker: Vionnie

Risk Management

4	4	Delay in completion of earlier phases of project increasing failure of project completion	Not enough time to meet the schedule target to complete the project	Resource risk/ Schedule Risk	Team member does not deliver task responsibilities on time	Time estimated for a certain task is not enough	Revise schedule estimates	Team	20%	Medium	Potential	
5	1	Predictive model has a low accuracy	The predictive model developed produces wrong outputs most of the time	Technical risk / Performance Risk	When users inputs a medical image to the predictive model	Predictive model not properly developed	Reidentify important predictors for the model, do more research on algorithms used in developing predictive model	Team	30%	High	Potential	
6	5	Losing source code	Source code is deleted and unable to be recovered	Technical risk	Source code accidentally deleted	Improper storage of project source code	Use GitHub for backup to minimise the risk	Team	5%	High	Potential	

Speaker: Vionnie

Risk Management

7	7	Website downtime	The platform that we decide to host our website in is unreliable and can't handle major traffics	Operation al Risk	Could be caused by massive traffic or overall unreliability	Not choosing a good platform to host the website	Researching on pros and cons of several website hosting platforms and choose the platform with the least risk	Team	10%	Medium	Potential	
8	10	Slow Stakeholder Actions that delays overall project completion	Poor communication with stakeholders and lack of verbal support from stakeholders		Attempting to have Communicative Measures with stakeholders but stakeholders remain unresponsive	Stakeholders are difficult to get hold of	Conduct regulatory meetings and emails to stakeholders. Have a stakeholder communication plan and update it accordingly. Make sure stakeholders are updated through every changes in the project	Team	5%	Low	Unlikely	
9	11	Scope Creep	Addition of unnecessary extra functionality not originally stated in the project scope	Scope Risk	Wanting to add new functionalities to enhance user experience but not addressing triple constraints of project	Excessive ideas given by team members	Clearly and succinctly state the requirements and scope in the project proposal. Update business case in case of changes	Project Manager / Team Members	5%	Low	Unlikely	

Speaker: Vionnie

Risk Management

10	12	Incomplete project design and deliverable definition	Not following preliminary designs created and risking to create a whole new environment that may have different deliverables	Technical Risk	Unclear preliminary designs that may not align with project scopes and definitions	Failure in understanding the given project and what is deemed an appropriate design and deliverable.	Create a preliminary design of the website, and build on it accordingly at each phase of the project.	Team	5%	Low	Unlikely	
11	2	Software does not fulfill customer requirements	Customers are collectively unable to obtain, view, save their results due to difficulty in accessing the user interface of the website	Software Risk	Customers may not be inclined to use the website for their purposes due to it being obsolete and not user-friendly	When customers are not satisfied with the product	Ensure that a substantial amount of testing is done on the website		15%	High	Potential	
12	3	Lack of coding capabilities leading to failure of the whole project	Lack of self-awareness in coding capabilities may result in our team abandoning the whole project	Technical risk	Lack of desire to understand the resources required to build the predictive model and website.	Lack of self-awareness on coding skills that may be exaggerated in order to	Being honest with current coding capabilities, and attending extra workshops/consultations both from Monash and online resources to ensure basic foundational knowledge.	Personal	15%	High	Potential	

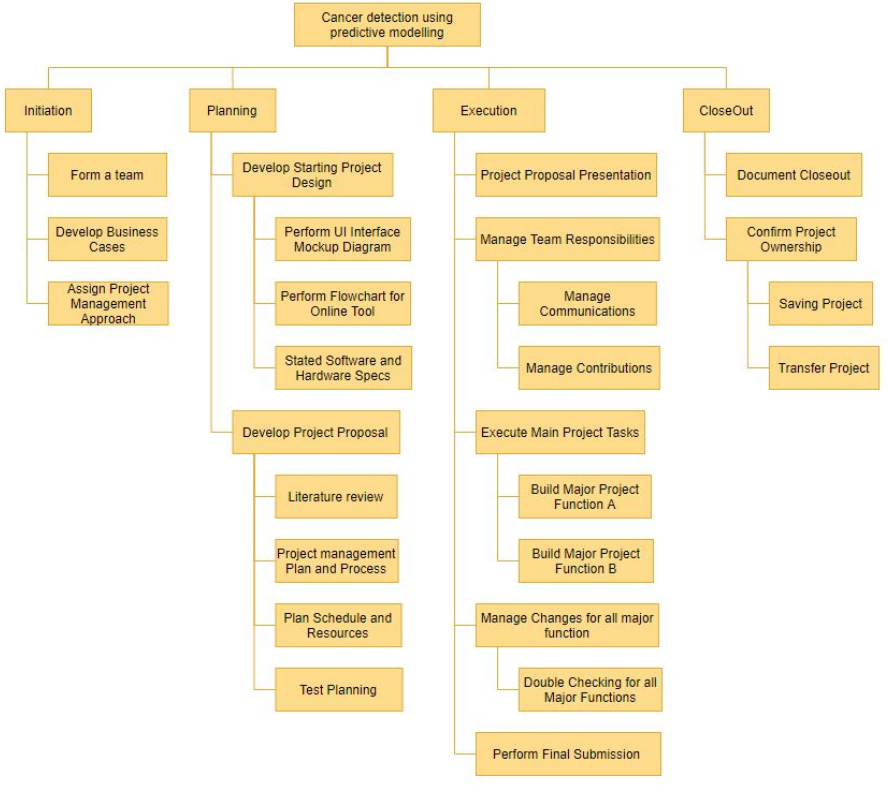
Speaker: Vionnie

Communication Plan

Stakeholders	Document Name	Document Format	Contact Person	Due
Project Manager	Progress report	Soft Copy via Google Doc and Zoom meetings	Afsaneh Koohestani	Weekly
Internal Management	Weekly status report	Soft copy and Zoom meetings	Jack Ooi, Elaine Liong, Vionnie Tan	End of week
Internal Management	Daily communication	Communication via Messenger	Jack Ooi, Elaine Liong, Vionnie Tan	Daily

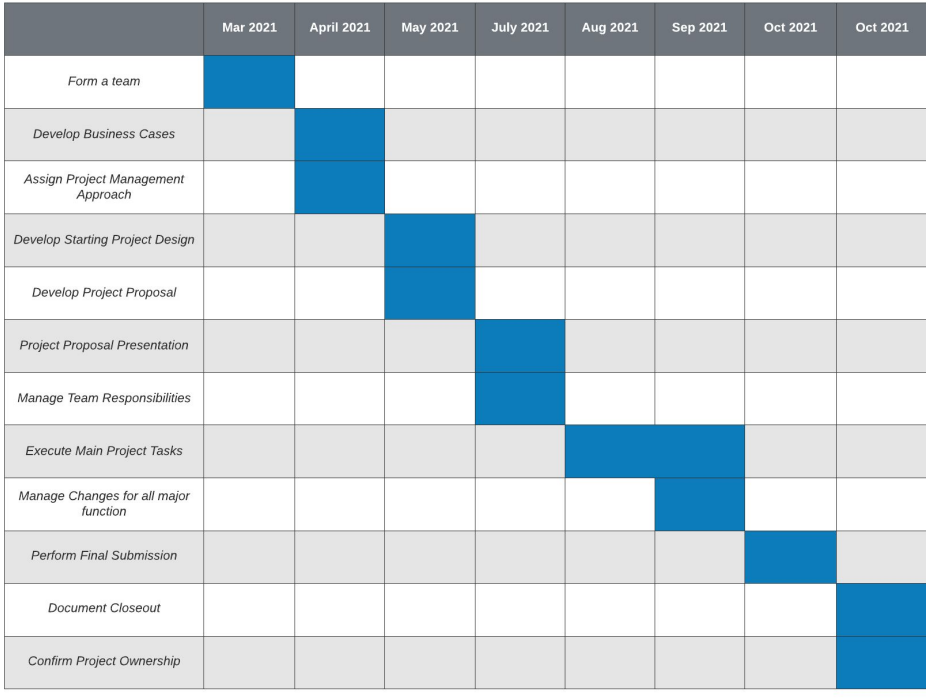
Speaker: Elaine

Schedule - WBS



Speaker: Elaine

Schedule - Gantt Chart

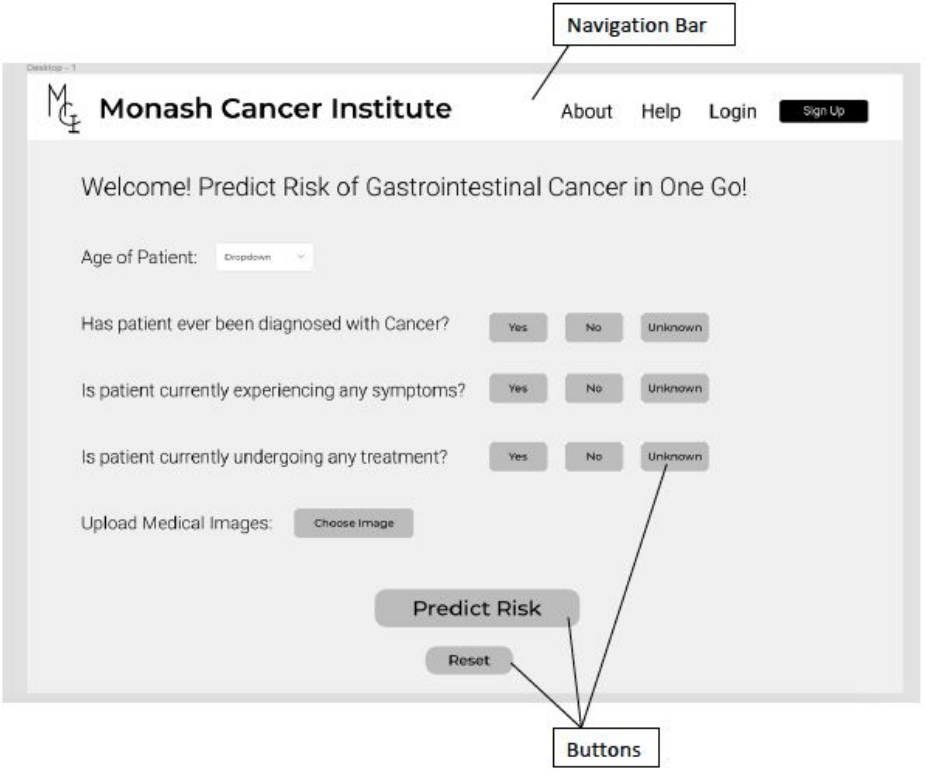


Speaker: Elaine

PROJECT DESIGN

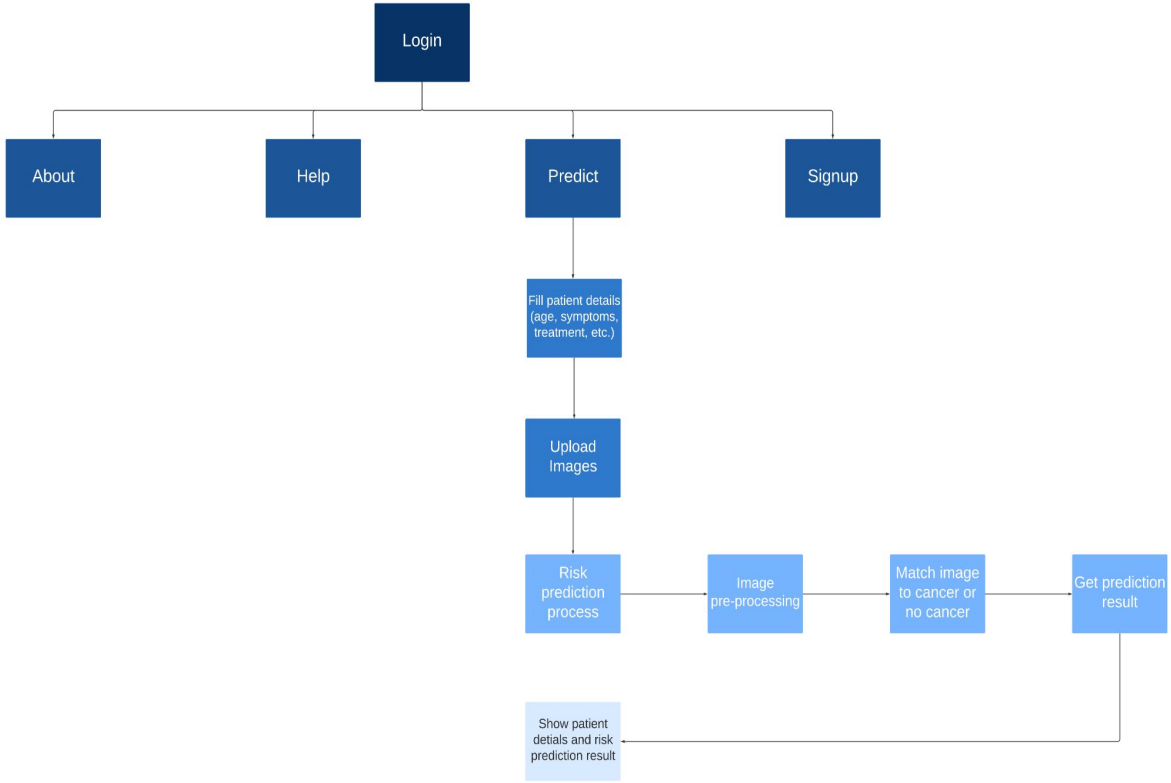
Speaker: Vionnie

User Interface



Speaker: Vionnie

Website Flowchart



Speaker: Vionnie

METHODOLOGY

Speaker: Elaine

Programming Language

Speaker: Elaine

Libraries

Speaker: Elaine

Libraries

Speaker: Elaine

Version Control System

Speaker: Elaine

Data Pre-Processing

- Dataset have been pre-processed
  - Images have automatic detection of tumor
  - Resized to 224 px x 224 px at a resolution of 0.5 µm/px
  - Color normalization using Macenko method
  - Patients are assigned to either "MSS" or "MSIMUT"
  - Reformatted to JPG format
- Feature extraction
- Divide dataset into 70% training set and 30% testing set

Speaker: Elaine

TEST PLANNING

Speaker: Vionnie

Predictive Model Testing

- Use 30% testing set to evaluate model's performance
- Determine optimal cut off levels for each images using ROC
  - obtain model performance using AUC
- Further evaluate performance using concordance index (c-index)
- Generate confusion matrix
  - Measure model accuracy, precision, recall, specificity

Speaker: Vionnie

Website Testing

- Handle mass traffic with reduced website downtime
- Providing same, accurate results as predictive model
- User’s private information is kept confidential

Speaker: Vionnie

CONCLUSION

Speaker: Jack

Summary

- Literature Review
  - History of Cancer Classification
  - Issues of Medical Data Mining
  - Classification Techniques
  - Image processing Techniques
- Project Management Plan
  - Project requirements
  - Project Organisation
  - Project Responsibilities
  - Risk Management
  - Communication Plan
  - Schedule
- Project Design
  - User Interface
  - Website Flowchart
- Methodology
  - Programming Language & Libraries
  - Version Control System
  - Data Pre-processing
- Test Planning
  - Predictive Model Testing
  - Website Testing

Speaker: Jack

REFERENCES

Speaker: Jack

References

Ansary, U, & Sarode, T. Skin Cancer Detection Using Image Processing.
Retrieved May, 12, 2021, from
https://d1wqtxts1xzle7.cloudfront.net/53555135/IRJET-V4I4702-with-cover-page.pdf?Expires=1620894864&Signature=DE~Uokb0e9tF5K6Dlv8FRj5kuA4u9nf0sz5q1vqpoYikD24WSZSI8SeWbFS7DPznq6vOoLolS2OG40dqqRhOgZVPV8LMtGQqL5YdRxbf5oMoFBurAhYpW22d6YEZuqI7JjdQtWlJcvfggeA4jchsWpO9kCtFuzyb-gQDVW5~6bJZxRHzGW4JvSALlBWPfm4DmrDxKp5S3iYpCqSuUWbHVNP5Z6P2qfrPVTl9VunTZ19abjwqMwTbupdvYtN5u9biJEpjW0fX0KdktcEpYwYRuEB4sLw0p9Z0xC5LOx3a4P1wF6CBVtsxaLCDiBM4msv8uw4VjLO6OhEEiC5Q4njuYw__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

Australian Institute of Health and Welfare. (2020, Nov 13). *Cancer data in Australia.* Rankings. Retrieved March 29, 2021, from
https://www.aihw.gov.au/reports/cancer/cancer-data-in-australia/contents/rankings

lom M, Z, Taha, T, M, Yakipcic, C & Westberg, S. (2018) The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches. Retrieved May 23, 2021 from
https://www.researchgate.net/publication/323570864_The_History_Began_from_AlexNet_A_Comprehensive_Survey_on_Deep_Learning_Approaches

Cancer Council Australia. (n.d.). *Cancer Statistics in Australia. Facts and Figures.* Retrieved March 29, 2021, from https://www.cancer.org.au/cancer-information/what-is-cancer/facts-and-figures

Delen, D. (2009). Analysis of cancer data: a data mining approach. Retrieved May 12, 2021
https://onlinelibrary.wiley.com/doi/full/10.1111/j.1468-0394.2008.00480.x

Speaker: Jack

References

Deng, Q., He, B., Liu, X., Yue, J., Ying, H., Pan, Y., Sun, H., Chen, J., Wang, F., Gao, T., Zhang, L., & Wang, S. (2015). Prognostic value of pre-operative inflammatory response biomarkers in gastric cancer patients and the construction of a predictive model. *Journal of Translational Medicine*, *13*(1), 66. https://doi.org/10.1186/s12967-015-0409-0

Demir, A & Yilmaz, F. (n.d). Early Detection of Skin Cancer Using Deep Learning Architectures: Resnet-101 and Inception-v3. Retrieved May 24, 2021. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8972045

Furey, T, S , Cristianini, N, Duffy, N, Bednarski, D,W, Schummer, M & Haussler, D. (2000). Support vector machine classification and validation of cancer tissues samples using microarray expression data. Retrieved May 24, 2021. https://watermark.silverchair.com/160906.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAArswggK3BgkqhkiG9w0BBwagggKoMIICpAIBADCCAp0GCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMGkOu07HHTGFA51ofAgEQgIICbr5TBibsaoxzqIqgib_ZvoNNTo_ODdXl4vyeTMxFZzzaLfF_1tvQdVtJX8WsZyW6_BLbKJOokHB-p5gmO2RwQAcAHwtT2dJ5h-ic9yL2X0eUQZyvFIVsMWS1hkuHTJZPVFzmSmeqYm08R4Q1TyGwBcQY-bxWx7h85Gdv0NZtMD6AU2mK8FmveleTJ984vnMU2BYad4gl1zfTr0UM4ycM3JWSX1XJo412rHeFs2DWaMUEC_J7KOyMlMQBQJvnf6RGRW_vOuMhhwPwttl8siZA6rlF17rbeJIK_rGhdFk6wARp83Tq4_vfQ94Z3PFAYLz2th5WG4_E2E5LrKwjbHweqaYPAwTKoVRp9fFn3o3MsY9NzDkQlSUIpE8NWZXzoECozxinp7xJN7egiLimm5_59ESUiA6tGsu1_IHIlsilY9acFUKYw9NMfvul6SwiPp3iltLVYFoyuBhQySTDZuXxCsOI_WDWti4frLpQcmP0LgB5cBGXba5UJAzQhy7uiV7E8o7cToAV18kwhyP5nzWNAN5rmd8N5zzMshX6lYPkOJPymtX1tFhyXtmUQORfCU219rn35zxtabzPbmyLQx37Xo5uWu-wA_FL5LDsh8N9G0tN44Ka8SVIlXJkUo2OgoNhxXRbNZJU0XLHxKWRncC5ZvvKEsVpZOn12BfdBgX_oi2ZJabV32NqNrzwae8DfLygQ1wroP2ttHMU8xYWTElNXB6FeYIAbj3ezOH1GPNcKrLfhAXpNiNE5s6bpwHt6q7vfY5y1FybFjgfSFdUxPaH_HqmLDKrz3D_jx48coy2t4PJyc6Lc2Q7QOvM1pYPEaI

Speaker: Vionnie

References

Gupta, S., Kumar, D., & Sharma, A. (2011). Data mining classification techniques applied for breast cancer diagnosis and prognosis. *Indian Journal of Computer Science and Engineering, 2*(2), 188-195. https://www.researchgate.net/publication/268054343_Data_mining_classification_techniques_applied_for_breast_cancer_diagnosis_and_prognosis

Indolia, S, Goswami A, K, Mishra S, P & Asopa, P. (2018). Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach. Retrieved May 24, 2021 from https://www.sciencedirect.com/science/article/pii/S1877050918308019#:~:text=Convolutional%20Neural%20Network%20(CNN)%20is,of%20traditional%20machine%20learning%20approaches.&text=This%20study%20provides%20the%20conceptual,common%20architectures%2C%20and%20learning%20algorithms.

Jain, S. Jagtap, V. & Pise, N. Computer Aided Melanoma Skin Cancer Detection Using Image Processing. Retrieved May, 12, 2021, from https://www.sciencedirect.com/science/article/pii/S1877050915007188

Kumar, S., Govardhan, A., & Srinivas, B. S. (2014). Data Mining Issues and Challenges in Healthcare Domain. *International Journal of Engineering Research, 3*(1), 857-861. https://www.ijert.org/research/data-mining-issues-and-challenges-in-healthcare-domain-IJERTV3IS10306.pdf

Speaker: Jack

References

Munir, K., Elahi, H., Ayub, A., Frezza, F., & Rizzi, A. (2019). Cancer Diagnosis Using Deep Learning: A Bibliographic Review. *Cancers, 11*(9), 1235.
https://res.mdpi.com/cancers/cancers-11-01235/article_deploy/cancers-11-01235.pdf

Serrador, P., & Pinto J. K. (2015). Does Agile Work? - A quantitative analysis of agile project success. *International Journal of Project Management, 33*(5), 1040-1051.
https://www.sciencedirect.com/science/article/pii/S0263786315000071

Speaker: Jack

THANK YOU!

Speaker: Jack