The current use of artificial intelligence in testicular cancer: a systematic review

Yanjinkham Chuluunbaatar¹, Saakshi Bansal¹, Andrew Brodie¹, Anand Sharma²,³, Nikhil Vasdev¹,⁴

¹Hertfordshire and Bedfordshire Urological Cancer Centre, Department of Urology, Lister Hospital, Stevenage SG1 4AB, UK.  
²Department of Oncology, Lister Hospital, Stevenage SG1 4AB, UK.  
³Department of Oncology, Mount Vernon Cancer Centre, Northwood HA6 2RN, UK.  
⁴School of Life and Medical Sciences, University of Hertfordshire, Hatfield AL10 9AB, UK.

Correspondence to: Prof. Nikhil Vasdev, Hertfordshire and Bedfordshire Urological Cancer Centre, Department of Urology, Lister Hospital, Corey Mill Lane, Stevenage SG1 4AB, UK. E-mail: nikhil.vasdev@nhs.net


Received: 2 Jul 2023 First Decision: 5 Sep 2023 Revised: 19 Sep 2023 Accepted: 7 Oct 2023 Published: 17 Oct 2023

Abstract

Testicular cancer is often overshadowed by other cancers despite being the most common cancer in men aged 15 to 34 years. This systematic review focuses on the potential of machine learning and deep learning techniques in the areas of testicular cancer imaging and histopathology, where artificial intelligence (AI) could assist in diagnosis, evaluation, and prognostication. Various studies have highlighted AI’s ability to accurately distinguish between benign and malignant lesions and characterisation within malignant lesions using magnetic resonance imaging (MRI) radiomics. Models have also been used in predicting histopathological findings to allow for greater accuracy and reproducibility. Further work is required to explore AI implementation in ultrasound imaging, which is the cheapest and most used modality.

Keywords: Artificial Intelligence, deep learning, machine learning, testicular cancer, oncology

INTRODUCTION

The term “artificial intelligence” (AI) was first coined for the Dartmouth Summer Workshop in 1956, where it was broadly referred to as “thinking machines”⁴. Artificial intelligence can be categorised into machine learning (ML) and deep learning (DL). Machine learning utilises statistical models to generate algorithms to
assess data and variables, referred to as “features”, to predict outcomes. With further data input, machine learning can improve its own performance over time.

The application of AI is vast and expanding rapidly in oncology[2] with the use of machine learning and deep learning. Interestingly, research as early as 1995 used deep learning neural networks to quantitate primary tumour to determine if more accurate staging could be achieved, which showed promising results[3]. Since then, early adoption of AI in urology has been beneficial, with a particular focus on renal, bladder, and prostate cancer[4,5]. However, testicular cancer is often overshadowed by the three aforementioned cancers, despite being the most common cancer in men aged 15 to 34 years[6], with annual incidence estimated to reach 23,000 in Europe by 2025, an increase of 24% from 2004[7,8].

**Testicular cancer**
The WHO classification of testicular cancer broadly distinguishes between germ cell neoplasia in situ (GCNIS; 95%) and non-GCNIS (5%), with the former further subdivided into seminomas and non-seminomas[6]. There are various important diagnostic values in diagnosing testicular cancer. Ultrasonography (US) is the most common first-line imaging modality in patients with suspected testicular cancer[7], but it is not able to confidently distinguish between benign and malignant testicular masses[9-11]. Computed tomography (CT) imaging plays a crucial role in the initial staging of disease prior to orchidectomy[12]. Magnetic resonance imaging (MRI) also provides greater resolution and diagnostic power over ultrasound[13,14] and can be combined with advanced image processing techniques to extract a large number of quantitative features and assist both prediction and diagnosis[15].

In conjunction with imaging, three important blood serum tumour markers carry diagnostic value at initial diagnosis and after orchidectomy - serum alpha-fetoprotein (AFP), beta subunit of human chorionic gonadotropin (beta-HCG), and lactate dehydrogenase (LDH). Elevation of these three tumour markers has been seen in up to 60% of patients[9]. Post-orchidectomy tumours markers, staging CT scan results, and histopathological results are used to classify patients in TMN (tumour, metastasis, and lymph node involvement) classification and their prognosis. Patients with lymphovascular invasion, typically involving retroperitoneal lymph nodes, have a higher risk of recurrence of cancer[16]. These are central to the use of AI in testicular cancer.

**Machine learning and deep learning**
Machine learning predominantly uses either supervised or unsupervised learning. Supervised learning is frequently employed by research teams when there is a clearly defined outcome, often of a binary nature, such as distinguishing between malignant and benign cases. The data requires meticulous annotation, a labour-intensive pre-processing endeavour, which prepares the dataset for the algorithm. Commonly used supervised learning ML algorithms include, but are not confined to, linear regression, random forest, and support vector machines (SVM). The concept of train-test split is fundamental to supervised machine learning; 70%-80% of the dataset is allocated to a training set and the remaining 20%-30% is used to validate the performance of the training model. In addition, cross-validation can be used and is most useful in smaller datasets and mitigates overfitting (an issue whereby machine learning does well in the training cohort but less well in unseen data in the testing cohort)[17,18].

On the other hand, unsupervised learning does not involve labelled output data and is used to identify commonalities of “features” within the dataset. In clinical practice, this type of learning is useful in identifying unrecognised patterns, clusters, or feature subsets, and commonly used algorithms include K-means (clustering) and principal component analysis[19]. Further to this, deep learning is a subset of machine learning that uses units called artificial neural networks (ANNs) with multiple layers that mimic
the human brain. The neural networks apply weights to different inputs, and pass these through a simple mathematical algorithm called an activation function. Output values are generated when data is filtered through these layers of networks. One commonly used DL model is convolutional neural network (CNN)\(^\text{[4,17,19]}\).

**METHODS**

This review was designed and performed in accordance with the recommendations of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines.

**Inclusion criteria**
- Original research full articles relevant to artificial intelligence including machine learning and deep learning in testicular cancer, seminomas, or germ cell tumours.
- Live human study participants

**Exclusion criteria**
- Grey literature
- Abstracts from conferences
- Non-peer-reviewed articles
- Non-English articles
- Animal studies
- Paediatric studies

**Literature search and study selection**
A systematic search of Cochrane CENTRAL database, MEDLINE, EMBASE, NHS NICE healthcare databases advanced search interface was executed by two authors (Chuluunbaatar Y and Bansal S). The last search was conducted on the 28th of June 2023. Medical subject headings (MeSH) terms were used - "artificial intelligence testicular cancer", "machine learning testicular cancer", and "artificial learning seminoma germ cell tumour". Two independent screeners performed a title and abstract review, with a third screener to resolve any disagreements (Brodie A).

**Data collection**
For each study included in this systematic review, the following data was extracted: the application, imaging technique where applicable, cancer type, sample size, training cohort size, validation cohort size, machine learning or deep learning AI algorithm, accuracy, sensitivity, specificity, and/or area under curve (AUC). Data extraction was performed independently by two authors (Chuluunbaatar Y, Bansal S), and disagreements were resolved by discussion and consensus. If no agreement could be reached, a third author (Brodie A) was consulted to act as an arbitrator. Studies with missing data were excluded from the systematic review.

**Assessment for risk of bias**
Two authors (Chuluunbaatar Y, Bansal S) critically appraised the methodological quality of the included studies. Discrepancies were resolved by reviewer discussion, and if they remained unsolved, a third investigator was consulted (Brodie A).

**Search result:** 102 citations were identified. Following screening and assessment, 9 studies were included in the review [Figure 1].
RESULTS

AI in testicular cancer imaging

Different imaging modalities are important in the diagnostic work-up and initial staging of testicular cancer. While ultrasound is essential in diagnosis, it cannot differentiate benign from malignant types\(^{20}\), and thus, limited work on AI and ultrasound has been carried out.

Recent studies involving MRI have been promising. Fan et al. developed a supervised machine learning-based approach to distinguish malignant and benign testicular masses using diffusion weight imaging (DWI) MRI “radiomic” signatures based on appearance diffusion coefficient (ADC) maps\(^{21}\). Radiomics refers to a widely used method of extracting quantitative features or data from digital medical images into high-dimensional mineable data\(^{22}\). In 97 patients with 101 testicular lesions, 851 distinct radiomic features were extracted from ADC maps. A predictive model was trained using a subset (71 lesions) and included 272 features, of which six were statistically significant and input into the AI algorithm. These performed well in a separate validation cohort [30 lesions and area under the curve (AUC) = 0.868] and, therefore, were able to distinguish malignant vs. benign masses well. The study also analysed blood tests, which
showed that levels of serum markers (AFP, hCG, and LDH) were increased in 67% of testicular malignancies in their training and validation cohorts. Further statistical analysis showed that it was not an independent factor for testicular malignancy.

Furthermore, Zhang et al. used minimum-redundancy maximum-relevance (mRMR) together with least absolute shrinkage and selection operator (LASSO), supervised machine learning algorithms, to design a classifier for seminomas and non-seminomas in 39 patients with germ-cell tumours using their T2-weighted MRI images prior to radical orchiectomy\cite{15}. As a labelled dataset, 19 patients were known to have seminomas and the remainder with non-seminomas. From these MRI images of 39 patients, a total of 851 radiomic features were extracted, of which features were ranked using the mRMR method according to their relevance-redundancy scores. Following this and 5-fold cross-validation, five features were included in the radiomic signature. The algorithms were able to statistically distinguish between seminomas and non-seminomas in this dataset (\(P < 0.01\)) and an AUC (area under the curve) of 0.979. While most germ cell tumours are managed with an orchiectomy, the distinguishing via AI between seminoma and non-seminoma could provide clinicians further guidance with regard to adjunct treatment.

Similarly, Feliciani et al. investigated the role of radiomics in 42 patients with histology-proven testicular neoplasms to differentiate between germ-cell tumours and non-germ cell tumours and between seminomas and non-seminomas within germ-cell tumours using T2-weighted MRI images prior to surgery\cite{20}. From 42 patients, 44 lesions were identified, and 487 features extracted. LASSO algorithm was applied and identified three radiomic features for the discrimination of germ-cell vs. non-germ cell tumours and four radiomic features for discrimination of seminoma vs. non-seminoma. These radiomic features were applied to a linear support vector machine (SVM) algorithm, a supervised machine learning algorithm. They showed that the model produced a true positive rate (TPR) of 94% and 74% in predicting germ-cell tumours and non-germ cell tumours, respectively. The F-score for germ-cell tumours is 0.92 and 0.78 for non-germ cell tumours, which is a score that provides a balanced measure of the model’s precision and recall. Within germ-cell tumours, the model was able to predict seminomas and non-seminomas with a TPR of 86% and 87%, respectively.

In addition to MRI, CT imaging has also been implemented in AI algorithms. Baessler et al. explored the potential for CT imaging to distinguish between benign and malignant lymph node involvement in 80 patients with metastatic non-seminomas who underwent post-chemotherapy retroperitoneal lymph node dissection (PC-RPLND)\cite{23}. They used gradient boost tree (GBT), a supervised machine learning algorithm, to develop a radiomics-based predictive model that achieved a classification accuracy of 0.81 compared to histological assessment. This study further solidified research carried out by Lewin et al. in which they used CT radiomics from 77 patients with PC-RPLND to distinguish residual masses as either malignant or benign (fibrotic) using supervised machine learning SVM\cite{24}. Ten radiomics features were identified and the algorithm performed well with an AUC of 0.80 compared with pathological reports.

These studies highlight the potential role AI can play in assisting diagnosis and classification via imaging. Further investigation is required to explore AI implementation in ultrasound imaging, the cheapest and commonly used modality.

**AI in testicular cancer histopathology**

Histological assessment by pathologists is prone to inter- and intra-observer bias. AI within histopathology has advanced exponentially and allows for greater accuracy and more reproducible results\cite{25,26}.
Tumour-infiltrating lymphocytes (TILs) have prognostic significance in testicular germ cell tumours such that their presence often indicates poorer prognosis\cite{27,28}, which are typically quantified histologically following haematoxylin and eosin (H&E) staining. Linder et al. utilised deep learning algorithms (commercially available but not specified) to assist in the detection of TILs in testicular germ cells in 259 specimens obtained retrospectively from 118 patients\cite{29}. The algorithm initially trained with 129 specimens, with the remaining 130 specimens reserved for testing, where the model performed with a sensitivity of 89% and F-score of 0.88 compared to manually annotated TILs. This showed high detection levels with good accuracy and precision.

In a similar fashion, Ghosh et al. carried out research on lymphovascular invasion (LVI) in 29 germ-cell testicular cancer patients retrospectively by using 184 whole slide images from primary management with orchidectomy retrospectively\cite{30}. The expertly annotated images were input into “Visiopham” AI module, which utilises deep learning convolutional neural network (CNN) for training. The model was applied to a test data set of 43 slides. The model demonstrated that it was able to identify lymphovascular invasion with an accuracy of 0.56. However, a few foci that the AI model identified as non-LVI were re-reviewed by an expert and determined to be in fact LVI. Although these models are not perfectly precise, they set the foundation for future work and highlight the importance of larger datasets required for stronger training and validation of AI models. It is particularly important as LVI is considered an important prognostic marker in non-seminomas, with a relapse rate of 50% if present\cite{31}.

**AI in testicular cancer outcomes**

A plethora of machine learning techniques have been devised to prognosticate patient outcomes, such as “DeepSurv”\cite{32}, “DeepHit”\cite{33}, and recently “Deep Survival Machine” models\cite{34}. However, these methods of estimating the survival distribution fail to provide insight into post-prognosis surveillance. To bridge this gap, Eminaga et al. used data from “Surveillance, Epidemiology and End Results (SEER) program” encompassing a vast cohort of cancer patients (penile, testicular, prostate, bladder, ureter, and kidney) to construct a machine learning model capable of dynamic risk-stratification for patient survival and follow-up\cite{35}. With a training set of 1.7 million patients, their model yielded a remarkable concordance index of 0.80 on the validation set (190,000 patients). Their results have also recommended at least five years of follow-up post-diagnosis for testicular cancer patients, and up to 8 years for localised cancer or 9 years for metastatic cases, thereby demonstrating the potential of AI in predicting survival and surveillance management. Additionally, they elucidated that the highest probability for a stabilised risk profile in testicular cancer with distant metastases is during the 5th and 7th year, emphasising the crucial role of AI in guiding surveillance management.

A small proportion of patients develop advanced disease stages characterised by metastases, typically to the lung followed by the liver\cite{36}. In a concerted effort to unravel the underlying risk factors and predict which patients are more susceptible to metastasis, machine learning algorithms are being harnessed. Recently, a study leveraged five supervised machine learning algorithms (linear regression, extreme gradient boosting, light gradient boosting, random forest, and k-nearest neighbours) and one deep learning algorithm (multilayer protection) to accurately prognosticate metastatic disease (M1B progression) in patients with GCTC. The study employed 4,300 patients from the SEER program database and evaluated various potential factors through statistical analysis\cite{37}. The results of the study revealed that T-stage, N-stage, lung metastases and distant lymph node metastases exerted the most statistically significant impact on M1B progression and therefore incorporated into the machine learning algorithm. XGBoost was identified as the most effective algorithm for forecasting metastatic risk, demonstrating an AUC of 0.957. Serum markers (AFP, LDH, and hCG) were also identified as not being a predictor of progression to metastatic stage. Despite the results, it is essential to note that immunohistochemistry findings were not factored into the algorithm, thereby leaving room for further valuable input into identifying risk factors.
DISCUSSION

This review has summarised recent novel research in the utilisation of AI in testicular cancer, as summarised in Table 1. As testicular cancer accounts for approximately 1% of all new cancer diagnoses in the world\[^6\], it is an uncommon cancer, and therefore, this is reflected in the small sample sizes of the vast majority of studies carried out. Consequently, many studies conducted feature small sample sizes. As a corollary, several studies lack a validation or testing cohort to test their algorithm [Table 1], and thus rely heavily on cross-validation. Multiple authors have acknowledged this as a common limitation, emphasizing the need for larger datasets to generate accurate predictions. This is particularly crucial in non-seminoma tumours, which are characterised by greater heterogeneity than seminomas, necessitating a greater number of samples to enable deep learning to assimilate a broader range of patterns.

The use of CT and MRI imaging radiomics is an emerging field in testicular cancer. One study used diffusion-weighted imaging MRI modality, while others used the commonly used T2-weighted MRI modality. Diffusion-weighted images are quantitatively assessed by calculating the ADC maps and scores, which offer information regarding cellular structure and organisation of tissues and, therefore, delineate between conditions, such that tissues with higher cell density, i.e., cancers, tend to have lower ADC values and vice versa\[^{38}\]. DWI, however, suffers from certain limitations such as geometric distortion and signal intensity dropout on tissue-air boundaries, such as thyroid gland, scrotum and prostate\[^{15}\]. Nevertheless, studies have demonstrated that DWI alone confers powerful diagnostic performance in testicular masse, with one study reporting sensitivity of 90% and specificity of 88%\[^{39}\]. Moreover, the amalgamation of ADC and conventional T2-weighed MRI has further improved diagnostics\[^{40,41}\]. The combination of the two has yet to be explored in AI algorithms and thereby an open scope for further investigation in this direction.

Despite a focus on CT and MRI radiomics, no current literature exists pertaining to the role of AI in ultrasound in testicular cancer, which is the first-line imaging modality. The use of AI in ultrasound has been predominantly explored in thyroid and breast, followed by cardiac, vascular, obstetrics and gynaecology, and musculoskeletal system\[^{42}\]. For instance, in the context of thyroid nodules, AI has been used to aid in the diagnosis of malignant and benign nodules, classify malignant modules, and prognosticate outcomes\[^{43}\]. In a study by Liu et al., a novel deep-learning CNN computer-aided model was employed to detect and classify thyroid nodules in ultrasound images\[^{44}\]. The deep learning model surpassed the diagnostic capabilities of radiologists as observed in sensitivity (0.964 vs. 0.928), specificity (0.780 vs. 0.366), and accuracy (0.928 vs. 0.816). Given both thyroid and testicular masses are investigated first line using ultrasound, these studies demonstrate the possible applications of US in testicular cancer. As such, this constitutes a critical area of investigation for future research.

Additionally, serum markers play a pivotal role in the diagnosis and prognostication of testicular, albeit as a non-independent marker\[^{7,37}\]. The emerging field of serum biomarkers and proteomics with machine learning is an exciting advancement and has been demonstrated in the prediction and prognosis of hepatocellular carcinoma (HCC)\[^{45}\]. Lee et al. demonstrated the use of machine learning with cell-free DNA (cfDNA), specifically AFP expression, detection system to establish a platform for HCC diagnosis and prognosis\[^{46}\]. Their approach showed improved accuracy in determining pathological features of HCC using a composite cfDNA score (which integrated total plasma cfDNA and cfAFP-DNA expression into a single score). Similar strategies could potentially be employed in the realm of testicular cancer with AFP, LDH, and hCG markers. Further research is warranted to explore the possibility of this.
<table>
<thead>
<tr>
<th>Study</th>
<th>Application</th>
<th>Study summary</th>
<th>Study type &amp; dates</th>
<th>Sample</th>
<th>Limitations</th>
<th>Algorithm/Model</th>
<th>Accuracy %</th>
<th>Sensitivity %</th>
<th>Specificity %</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fan et al., 2022[21]</td>
<td>Detection</td>
<td>Develop machine learning-based radiomics to discriminate between benign and malignant testicular masses</td>
<td>Retrospective single-site study from February 2014 to April 2021</td>
<td>97 patients with 101 testicular masses Extracted = 851 radiomics 272 statistically different 6 features used</td>
<td>Small training and validation cohorts</td>
<td>Supervised machine learning</td>
<td>90.5</td>
<td>66.7</td>
<td>86.8</td>
<td></td>
</tr>
<tr>
<td>Zhang et al., 2019[15]</td>
<td>Imaging</td>
<td>T2-weighted image-based radiomics for differentiating between seminoma and non-seminoma</td>
<td>Retrospective single-site study from February 2014 to March 2019</td>
<td>39 patients Pathologically identified lesions Five features selected as radiomic signatures</td>
<td>There is no training or validation cohort due to the small sample size. Instead, they used cross-validation</td>
<td>Supervised machine learning mRMR LASSO</td>
<td>NR</td>
<td>90</td>
<td>100</td>
<td>0.979</td>
</tr>
<tr>
<td>Feliciani et al., 2021[20]</td>
<td>Imaging</td>
<td>Using MRI (T2-weighted) based radiomics to distinguish between germ-cell tumours vs. non-germ cell tumours To distinguish between seminoma and non-seminoma cancers Against known histological findings</td>
<td>Retrospective single-site study between January 2006 and February 2019</td>
<td>42 patients 44 lesions 487 features extracted 3 radiomic features used in germ-cell vs. non-germ cell tumours 4 radiomic features used in seminoma vs. non-seminoma cancers</td>
<td>No training or validation cohort due to the small sample size. Cross-validation used to mitigate this</td>
<td>Supervised machine learning Linear model SVM</td>
<td>89%</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Baessler et al., 2020[23]</td>
<td>Imaging</td>
<td>CT radiomics-based machine learning to predict histopathology of LN after dissection in metastatic non-seminoma</td>
<td>Retrospective single-site study between 2008 and 2017</td>
<td>80 patients with 204 lymph nodes 97 radiomics features per lymph node. 5 features used in AI algorithm Training dataset = 120 LNs Testing dataset = 23 LNs Validation dataset = 61 LNs</td>
<td>Various CT scanners used Did not include clinical variables in the AI algorithm</td>
<td>Supervised machine learning GBT</td>
<td>0.81</td>
<td>88</td>
<td>72</td>
<td>NR</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Year</td>
<td>Section</td>
<td>Description</td>
<td>Study Details</td>
<td>Methodology</td>
<td>Accuracy</td>
<td>Specificity</td>
<td>Sensitivity</td>
<td>F-score</td>
<td>AI/Machine Learning</td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
<td>---------</td>
<td>-------------</td>
<td>---------------</td>
<td>-------------</td>
<td>---------</td>
<td>------------</td>
<td>------------</td>
<td>--------</td>
<td>------------------</td>
</tr>
<tr>
<td>Lewin et al., 2018</td>
<td></td>
<td>Imaging</td>
<td>Imaging Contrast CT radiomics in predicting post-chemotherapy residual mass and compared with pathological reports</td>
<td>Retrospective single-site study between January 77 patients 183 radiomics 10 statistical 302 whole haematoxylin and eosin slides</td>
<td>Supervised machine learning</td>
<td>72</td>
<td>56</td>
<td>82</td>
<td>0.80</td>
<td>Support vector machine</td>
</tr>
<tr>
<td>Ghosh et al., 2021</td>
<td></td>
<td>Histopathology</td>
<td>Detection of lymphovascular invasion in testicular cancer</td>
<td>Retrospective single-site study from January 2019 to July 2020 29 patients 184 whole slides from 19 patients. Validation set = 118 whole slides from 10 patients</td>
<td>Deep learning</td>
<td>0.68</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>Convolutional neural network</td>
</tr>
<tr>
<td>Linder et al., 2019</td>
<td></td>
<td>Histopathology</td>
<td>Detecting tumour-infiltrating lymphocytes from H&amp;E slides</td>
<td>Retrospective single-site study between 2005 and 2016 89 patients 28 whole slide H&amp;E tissue samples 259 images created</td>
<td>Deep learning</td>
<td>NR</td>
<td>89</td>
<td>NR</td>
<td>F-score = 0.88</td>
<td>Commercially available image analysis software WebMicroscope</td>
</tr>
<tr>
<td>Eminaga et al., 2022</td>
<td></td>
<td>Outcomes</td>
<td>Prognostic and follow-up recommendations</td>
<td>Retrospective study utilising a US SEER National database with 1.9 million patients with urological cancers between 1975 and 2017 1.9 million patients</td>
<td>Deep learning</td>
<td>0.8</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>Recurrent neural network</td>
</tr>
<tr>
<td>Ding et al., 2022</td>
<td></td>
<td>Outcomes</td>
<td>Predicting the risk of developing metastatic disease with germ-cell testicular cancer</td>
<td>Retrospective study US SEER national database Total 4,323 patients</td>
<td>Supervised machine learning</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>0.816</td>
<td>Multiple</td>
</tr>
</tbody>
</table>

AI: Artificial intelligence; AUC: area under curve; CT: computed tomography; DWI: diffusion weighted imaging; GBT: gradient boosting tree; H&E: haematoxylin and eosin; LASSO: least absolute shrinkage and selection operator; LNs: lymph nodes; NR: not reported; MRI: magnetic resonance imaging; SEER: Surveillance, Epidemiology and End Results; SVM: support vector machine; US: ultrasonography.
There are several limitations of the papers and data included worth noting. Primarily, these limitations stem from the novelty of using AI algorithms within research such that the studies incorporated in this systematic review carried out their work retrospectively. This subsequently means that primarily internal validation processes are employed. However, for a more comprehensive assessment of these AI algorithms, external validation is imperative following testing with larger datasets to enable a better understanding of their performance in real-world scenarios with unseen data in prospective patients. Additionally, the retrospective nature of the studies often necessitated the resampling of images in order to address the heterogeneity of images to create a more uniform dataset. However, it is worth noting that this process may not perfectly mirror the diverse range of imaging equipment used across different clinical centres.

In essence, while these limitations are inherent to the current state of AI research and its vastly developing field, they provide invaluable insights into the challenges and opportunities for further development and validation of promising AI technologies within a clinical context.

**CONCLUSION**

The diverse array of studies employing AI techniques to address various aspects of testicular cancer underscores the extensive scope and versatility of AI in facilitating diagnosis, classification, and prognostication, ultimately impacting patient care. However, it is worth noting that a significant challenge associated with machine learning is its dependence on large-scale training and testing datasets to ensure the creation of robust and generalisable models.

**DECLARATIONS**

**Acknowledgements**

Acknowledgement to Roy Zhang for reading the initial manuscript.

**Authors’ contributions**

Made substantial contributions to the conception and design of the study and performed data analysis and interpretation: Chuluunbaatar Y

Performed data acquisition, as well as providing administrative, technical, and material support: Bansal S, Brodie A, Sharma A, Vasdev N

**Availability of data and materials**

Not applicable.

**Financial support and sponsorship**

None.

**Conflicts of interest**

All authors declared that there are no conflicts of interest.

**Ethical approval and consent to participate.**

Not applicable.

**Consent for publication**

Not applicable.
REFERENCES


22. Society of Radiology (ESR). What the radiologist should know about artificial intelligence - an ESR white paper. Insights Imaging 2019;10:44. DOI PubMed PMC


42. Shen YT, Chen L, Yue WW, XuHX. Artificial intelligence in ultrasound. Eur J Radiol 2021;139:109717. DOI PubMed
45. Moldogazieva NT, Mokhosoev IM, Zavadskiy SP, Terentiev AA. Proteomic profiling and artificial intelligence for hepatocellular carcinoma translational medicine. Biomedicines 2021;9:159. DOI PubMed PMC