Risk Estimation of Knee Osteoarthritis Progression via Predictive Multi-task Modelling from Efficient Diffusion Model using X-ray Images

David Butler*, Adrian Hilton, and Gustavo Carneiro

Centre for Vision, Speech and Signal Processing, University of Surrey
{d.butler,a.hilton,g.carneiro}@surrey.ac.uk

Abstract. Medical imaging plays a crucial role in assessing knee osteoarthritis (OA) risk by enabling early detection and disease monitoring. Recent machine learning methods have improved risk estimation (i.e., predicting the likelihood of disease progression) and predictive modelling (i.e., the forecasting of future outcomes based on current data) using medical images, but clinical adoption remains limited due to their lack of interpretability. Existing approaches that generate future images for risk estimation are complex and impractical. Additionally, previous methods fail to localize anatomical knee landmarks, limiting interpretability. We address these gaps with a new interpretable machine learning method to estimate the risk of knee OA progression via multi-task predictive modelling that classifies future knee OA severity and predicts anatomical knee landmarks from efficiently generated high-quality future images. Such image generation is achieved by leveraging a diffusion model in a class-conditioned latent space to forecast disease progression, offering a visual representation of how particular health conditions may evolve. Applied to the Osteoarthritis Initiative dataset, our approach improves the state-of-the-art (SOTA) by 2%, achieving an AUC of 0.71 in predicting knee OA progression while offering 9× faster inference time.

Keywords: Risk estimation · Knee osteoarthritis · Predictive Multi-task Modelling · X-ray.

1 Introduction

Knee osteoarthritis (OA) is a degenerative joint disease characterised by cartilage breakdown, bone remodelling, and joint inflammation [25]. It is a leading cause of disability in older adults, resulting in pain, stiffness, and reduced function. The Kellgren-Lawrence (KL) scale is commonly used to grade osteoarthritis severity, ranging from 0 to 4 based on joint space narrowing, osteophytes, sclerosis, and bone remodelling [13], as shown in Fig. 1. Early diagnosis enables treatment to alter the disease course [3].

Medical imaging plays a central role in knee osteoarthritis (OA) risk estimation [5] by analysing tissue changes over time. Machine learning techniques

* Corresponding author.

Fig. 1. (Left) Example of a 0 KL grade. (Right) Example of a 4 KL grade with osteophytes (red), sclerosis (blue), and bone remodelling (green).

compute the likelihood of disease progression [27, 12, 28, 26, 4], but most methods generate only numerical scores, offering little visual explanation for clinicians [21]. For instance, if a model predicts OA progression based on X-rays, it is crucial to understand which features, such as OA severity or anatomical landmarks, contribute to this prediction. Predictive modelling has been rarely explored, except for [9], which employed a highly complex image generation process, limiting clinical practicality and lacking anatomical landmark localization. Combining predictive modelling with future image generation and anatomical landmark detection enhances interpretability, fosters trust, and supports informed decision-making.

This paper presents a new interpretable multi-task machine learning method for estimating the risk of knee OA progression by predicting future OA severity grade and anatomical knee landmark localisation from efficiently generated future images. Such image generation leverages an efficient diffusion model using a class-conditioned latent space to forecast disease progression, offering a visual representation of how such particular health conditions may evolve. Our key contributions include:

- A new interpretable machine learning method for knee OA risk estimation via multi-task prediction modelling for KL classification and anatomical knee landmark localisation using future images generated by a diffusion model;
- A novel, compact, and efficient diffusion model that can generate high quality future OA X-ray images conditioned only by current images.

Experiments show that our proposed method has state-of-the-art (SOTA) results on the Osteoarthritis Initiative (OAI) dataset [5], a study on knee osteoarthritis, delivering superior risk estimation AUC of 0.71 while being $\sim 9\times$ faster at inference than the previous SOTA[9] that has 0.69 AUC.

2 Related Work

Risk Estimation and Predictive Modelling methods assess risk by predicting clinical events [27, 12, 28, 26, 4] or forecasting future features [14, 17, 16, 20, 2]. While event prediction is useful, it lacks interpretability, as it does not explain underlying causes. For instance, multiple plausible progression pathways

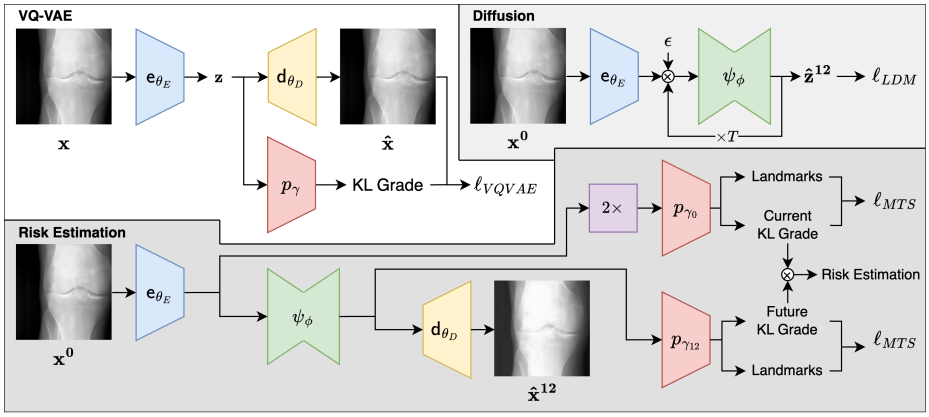

Fig. 2. Overview of the method. (Top Left) VQ-VAE training. (Top Right) Diffusion model training. (Bottom) Classifier training & inference with predicted future image $\hat{\mathbf{x}}^{\mathbf{12}}$ and the risk estimated from the KL grades predicted by p_{γ_0} and $p_{\gamma_{12}}$.

could lead to mortality, yet these models often do not differentiate between them. Similarly, feature prediction models estimate disease onset [14, 17, 16, 20] or severity [2], often using biomarkers [20, 18] and imaging data [18, 17]. However, their opaque reasoning limits clinical adoption [21].

Future image synthesis methods use StyleGAN [9, 19, 1], VAEs [10], flow-based models [15, 24], and diffusion models [30]. Some rely on an input image and patient information [19, 1, 10, 15, 24, 7, 9, 11], while others omit non-image data like biomarkers [15, 24, 7, 9, 11]. Diffusion models now surpass GANs in image quality [25] but remain computationally demanding and underutilized for disease progression risk estimation [30]. In knee OA research, StyleGAN has achieved SOTA accuracy [9], yet diffusion models offer superior image quality [25]. However, [9] does not generate anatomical knee landmarks, limiting interpretability.

3 Methodology

Let $\mathcal{D} = \{\mathbf{x}_i^0, \mathbf{x}_i^{12}, \mathbf{y}_i^0, \mathbf{y}_i^{12}, \{\mathbf{l}_{i,j}\}_{j=1}^{L}\}_{i=1}^{|\mathcal{D}|}$ represent the OAI dataset, where $\mathbf{x}^0, \mathbf{x}^{12} \in \mathcal{X} \subset \mathbb{R}^{H \times W}$ are knee X-ray images of a patient at an arbitrary point in time, and 12 months afterwards, respectively. Corresponding one-hot 5-class KL classifications are $\mathbf{y}^0, \mathbf{y}^{12} \in \mathcal{Y} \subset \{0, 1\}^5$. The set of L anatomical knee landmarks at $\mathbf{x}^0$ is $\{\mathbf{l}_{i,j}\}_{j=1}^{L} \in \mathcal{L}$, with each landmark $\mathbf{l}_{i,j} \in \{1, \ldots, H\} \times \{1, \ldots, W\}$. Our model comprises: 1) VQ-VAE for latent image generation, 2) a conditional diffusion model for future latent images, and 3) a multi-task classifier for OA severity prediction and anatomical knee landmarks localization (Fig. 2).

VQ-VAE: Future image generation for risk estimation leverages diffusion models, which perform better in latent spaces than in image spaces [22]. To generate this latent space, we use VQ-VAE, as it offers superior reconstruction quality and efficiency compared to VQ-GAN [6]. VQ-VAE consists of an encoder $\mathsf{e}_{\theta_E} : \mathcal{X} \to \mathcal{Z}$ and decoder $\mathsf{d}_{\theta_D} : \mathcal{Z} \to \mathcal{X}$, with $\mathcal{Z} \subset \mathbb{R}^Z$ as the latent space, parameterised by $\theta = \{\theta_E, \theta_D\} \in \Theta$. Following [22], we enhance perceptual quality and classification by integrating a classifier $p_\gamma : \mathcal{Z} \to \Delta^4$ for 5-class KL classification, forming a multi-task autoencoder [8]. The model is trained with:

$$\ell_{VQVAE}(\theta, \gamma) = \mathbb{E}_{\mathbf{x},\mathbf{y}\sim\mathcal{D}}\Big[\log(p(\mathbf{x}|\mathbf{z}_q(\mathbf{x}))) + ||sg(\mathbf{z}_e(\mathbf{x})) - \mathbf{e}||_2^2 + \beta||\mathbf{z}_e(\mathbf{x}) - sg(\mathbf{e})||_2^2 - \alpha \sum \mathbf{y}^T \log(p_\gamma(\mathbf{z}_e(\mathbf{x})))\Big], \tag{1}$$

where $\mathbf{x}$ is the input image, $\mathbf{z}_e(\mathbf{x}) = \mathsf{e}_{\theta_E}(\mathbf{x})$ is its embedding, $\mathbf{z}_q(\mathbf{x})$ the quantised embedding, $sg(.)$ the stop-gradient operator, $\mathbf{e}$ the nearest codebook entry, β controls adherence to the nearest codebook entry, α weights the classification term, $\mathbf{y}$ is the one-hot class label, and $p_\gamma(.)$ the classifier operating in the latent space of the diffusion model. This approach improves the classification accuracy of future synthetic images generated by the diffusion model.

Conditional Diffusion Model: The conditional diffusion model $\mathsf{g}_\phi : \mathcal{Z} \to \mathcal{Z}$, parametrised by $\phi \in \Phi$, generates future image embeddings (12 months ahead) conditioned on a patient's current embedding in the latent space $\mathcal{Z}$. Following [22], it learns $\mathsf{g}_\phi(\mathbf{z})$ by iteratively denoising Gaussian noise $\epsilon \sim N(0, I)$, using a U-Net with $\mathbf{v}$-prediction [23], minimising:

$$\ell_{LDM}(\phi) = \mathbb{E}_{\boldsymbol{\epsilon},\mathbf{z}^{12},t,\mathbf{z}^0}\left[||\mathbf{v} - \mathbf{v}_\phi(\mathbf{z}_t^{12}, t, \mathbf{z}^0)||_2^2\right], \tag{2}$$

where $\mathbf{v} = \alpha_t\epsilon - \sigma_t\mathbf{z}^{12}$ is a velocity vector, with α_t and σ_t denoting noise and signal proportions at step t, $\mathbf{v}_\phi$ is estimated via U-Net, $\mathbf{z}_t^{12}$ is the latent embedding of the future image, and $\mathbf{z}^0$ represents the conditioning image embedding, concatenated with $\mathbf{z}_t^{12}$ for conditioning. The U-Net has four encoding/decoding blocks and a bottleneck, with spatial self-attention in the first three and last three blocks, and channel-wise attention elsewhere. Inference model weights are obtained through an exponential moving average during training.

Risk Estimation via Predictive Modelling: Risk estimation uses the conditional diffusion model $\mathsf{g}_\phi(\mathbf{z}^0)$ to generate the future embedding $\hat{\mathbf{z}}^{12}$ from current image embedding $\mathbf{z}^0$. Two classifiers, denoted by $p_{\gamma_0} : \mathcal{Z} \to \Delta^4$ and $p_{\gamma_{12}} : \mathcal{Z} \to \Delta^4$, independently classify both $\mathbf{z}^0$ and $\hat{\mathbf{z}}^{12}$. The risk, defined as the probability of an increase in KL grade between $\mathbf{z}^0$ and $\hat{\mathbf{z}}^{12}$ [9], is computed as:

$$p(y = 1 \mid \mathbf{z}^0, \hat{\mathbf{z}}^{12}) = \sum_{c<k} p_{\gamma_0}(y^0 = c \mid \mathbf{z}^0) \cdot p_{\gamma_{12}}(y^{12} = k \mid \hat{\mathbf{z}}^{12}), \tag{3}$$

$$p(y = 0 \mid \mathbf{z}^0, \hat{\mathbf{z}}^{12}) = \sum_{c\geq k} p_{\gamma_0}(y^0 = c \mid \mathbf{z}^0) \cdot p_{\gamma_{12}}(y^{12} = k \mid \hat{\mathbf{z}}^{12}), \tag{4}$$

where $y = 1$ indicates an increase in KL grade, $y = 0$ indicates no increase, y^0 is the current KL grade, y^{12} is the KL grade after 12 months, and $c, k \in \{0, 1, 2, 3, 4\}$ iterate over KL grades. The classifier from VQ-VAE multi-task learning serves as an initial model for fine-tuning risk estimation, using

$$\ell_{CLS}(\gamma_0, \gamma_{12}) = \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{D}}\left[-\mathbf{y}^T \log\left(p_\gamma\left(\mathbf{y}|\mathbf{z}(\mathbf{x})\right)\right)\right], \tag{5}$$

where γ_0 is estimated from $\mathbf{x}^0, \mathbf{y}^0$, and γ_{12} from $\mathbf{x}^{12}$ and $\mathbf{y}^{12}$, both in $\mathcal{D}$. Moreover, $\mathbf{z}^0$ can optionally be upscaled $2\times$ with bicubic interpolation at test time, as shown in Fig. 2 – we note in the experiments of Sec. 4.3 that such upscaling enables more accurate predictions.

Multi-task learning The multi-task classifier improves classification while predicting anatomical knee landmarks for interpretation. It is defined as $p_\zeta : \mathcal{Z} \rightarrow \Delta^4 \times \mathcal{L}$, where $\mathcal{L}$ represents L knee landmark coordinates. Deconvolutional layers are added to the classifier, followed by a 2D SoftArgmax function [29]. The model is trained using:

$$\ell_{MTS}(\zeta) = \mathbb{E}_{(\mathbf{x},\mathbf{y},\{\mathbf{l}_j\}_{j=1}^L)\sim\mathcal{D}}\left[-\mathbf{y}^T \log\left[p_\zeta\left(\mathbf{y}|\mathbf{z}(\mathbf{x})\right)\right] + \delta\sum_{j=1}^{L}||\mathbf{l}_j - \hat{\mathbf{l}}_j||_2^2\right], \tag{6}$$

where $\mathbf{y}$ is the true KL grade for latent image embedding $\mathbf{z}_e(\mathbf{x})$, $\mathbf{l}_j = [x_j, y_j]$ is a 2-dimensional landmark coordinate, $\hat{\mathbf{l}}$ is the model's prediction, and δ is a weighting hyperparameter.

Training Algorithm Training starts by optimizing VQVAE and its classifier, $p_\gamma(.)$ with ℓ_{VQVAE} in Eq. (1). The trained VQVAE works as the foundation for training the latent diffusion model, $\mathbf{g}_\phi(.)$, with the loss ℓ_{LDM} in Eq. (2). Once trained, the latent diffusion model generates future X-ray images for all dataset samples. Next, classifiers $p_{\gamma_0}(.)$ and $p_{\gamma_{12}}(.)$ are fine-tuned from $p_\gamma(.)$ using ℓ_{CLS} in Eq. (5), leveraging ground truth and generated future images, respectively. Alternatively, these classifiers can be optimized with ℓ_{MTS} in Eq. (6) to jointly learn KL classification and anatomical knee landmark prediction.

4 Experiments

4.1 Dataset and Assessment

The Osteoarthritis Initiative (OAI) dataset contains 47,027 knee radiographs from 4,796 patients [5], captured at 0-, 12-, 24-, 36-, 48-, 72-, and 96-month intervals. Each image is KL-graded, excluding total knee replacements, which cannot be classified. Landmark coordinates for $L = 16$ joint surface points are provided for 748 images. Following [29], all images are cropped to 512^2 pixels using a landmark prediction model, ensuring full knee visibility. Left knee images

are flipped for consistency. The dataset is split into training (3,772), validation (512), and testing (512) patients.

Evaluation spans classification, prediction, and risk estimation. Classification involves estimating the current KL class $y^0 \in \{0, 1, 2, 3, 4\}$ from $\mathbf{x}^0$ or a latent representation $\mathbf{z}^0$. Prediction forecasts KL class y^{12} 12 months ahead. Risk estimation generates a future latent image $\hat{\mathbf{z}}^{12}$ from $\mathbf{z}^0$ using the conditional diffusion model, predicts the KL classifications y^0 and y^{12}, and calculates the binary probability of KL class progression over 12 months based on Eqs. 3 and 4. Decreases in ground-truth KL class from y^0 to y^{12} are assumed to be noisy and are instead treated as stable.

Classification and prediction performance is measured using the mean area under the receiver operating characteristic curve (mAUC), computed as the average of AUC values for each class in a one-vs-rest manner. Risk estimation is simply measured with the AUC. We compare our method to [9], the current SOTA for risk estimation via image generation for knee OA.

4.2 Training

The **VQ-VAE** is trained on the training fold for 5 epochs with a mini-batch size of 8. It uses an Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$) and a cosine scheduler (initial LR 10^{-4}, minimum LR 10^{-6}). Multi-task training with classification uses $\alpha = 10^{-4}$. The model has a compression ratio of 8, a codebook size of 256, and integrates vector quantization with the decoder.

The **conditioned diffusion model** is trained on image pairs spaced 12 months apart: {0,12}, {12,24}, {24,36}, and {36,48}. Images from 72 and 96 months are excluded due to 24-month gaps. Training runs for 200 epochs with a mini-batch size of 8, using an Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.99$) and a cosine scheduler (initial LR 10^{-4}, minimum 10^{-6}). The diffusion process uses 1000 time steps, and sampling applies an exponential moving average of weights with $\gamma = 0.995$.

The **classifier** is trained on true 0-, 12-, 24-, and 36-month images, and the second classifier is trained on synthetic 12-, 24-, 36-, and 48- month images generated by the diffusion model with 100 time steps for faster inference. Training uses mini-batches of size 8, balanced by whether KL progression occurs. Multi-task classifiers estimates anatomical knee landmarks, trained similarly with a landmark loss weight $\delta = 0.5$.

4.3 Ablation Study

Classification: Tab. 1 shows lower performance in latent space than image space. However, training the classifier within VQ-VAE mitigates this drop, and fine-tuning further improves results, surpassing image-space classification.

Prediction: Tab. 2 shows lower accuracy than classification (Tab. 1) since labels are not directly derived from input images. Latent-space prediction underperforms compared to image space, but training the classifier in VQ-VAE improves

Table 1. Ablation study on classification.

Experiment	mAUC
Image space $p(y^0\|\mathbf{x}^0)$	0.82
Latent space $p(y^0\|\mathbf{z}^0)$	0.66
+ VQ-VAE classifier training	0.70
+ fine-tune VQ-VAE classifier	0.87

Table 2. Ablation study on prediction.

Experiment	mAUC
Image space $p(y^{12}\|\mathbf{x}^0)$	0.80
Latent space $p(y^{12}\|\mathbf{z}^0)$	0.57
+ VQ-VAE classifier training	0.71
+ fine-tune VQ-VAE classifier	0.84

Table 3. Ablation study on risk estimation.

Experiment	AUC
Image space (ground truth $\mathbf{x}^{12}$) $p(y^{12} > y^0\|\mathbf{x}^0, \mathbf{x}^{12})$	0.75
Latent space $p(y^{12} > y^0\|\mathbf{z}^0, \hat{\mathbf{z}}^{12})$	0.57
+ VQ-VAE classifier training	0.60
+ fine-tune VQ-VAE classifier	0.63
+ multi-task training (classifier+landmark localisation)	0.65
+ 2 × upscale $\mathbf{z}^0$	0.71

results, with fine-tuning further enhancing performance. Despite achieving a high mAUC of 0.84, this method predicts only probabilities, making interpretation difficult, and remains less complex than risk estimation, which requires accurate predictions of both y^{12} and y^0, as discussed in the next section.

Risk Estimation: Tab. 3 evaluates risk estimation, $p(y^{12} > y^0|\mathbf{z}^0, \hat{\mathbf{z}}^{12})$. The diffusion model generates $\hat{\mathbf{z}}^{12}$, but image-space evaluation using $\mathbf{x}^0, \mathbf{x}^{12}$ is also considered for reference. Latent-space performance is lower since the image-space evaluation benefits from the ground truth future images. Training the classifier in VQ-VAE improves results, further enhanced by fine-tuning and multi-task learning with landmark prediction. Upscaling $\mathbf{z}^0$ at test time significantly boosts performance with an AUC of 0.71, corresponding with a sensitivity of 0.83, specificity of 0.51 and F1 score of 0.55.

4.4 Comparison with SOTA

Our method surpasses SOTA in OAI risk estimation (AUC 0.71 vs. 0.69 [9]) with significantly higher efficiency. Our training takes 12.6 hours on a single Nvidia A6000, compared to 114.88 hours on 2× A6000s for [9], while our inference is 8.7× faster (2.70s vs. 23.6s per sample). Additionally, our approach improves interpretability by not only generating future images but also localizing anatomical knee landmarks, as illustrated in Fig. 3. Beyond generating images that better align with ground truth and providing landmark estimations, our method produces higher-resolution images than [9], further enhancing result interpretability.

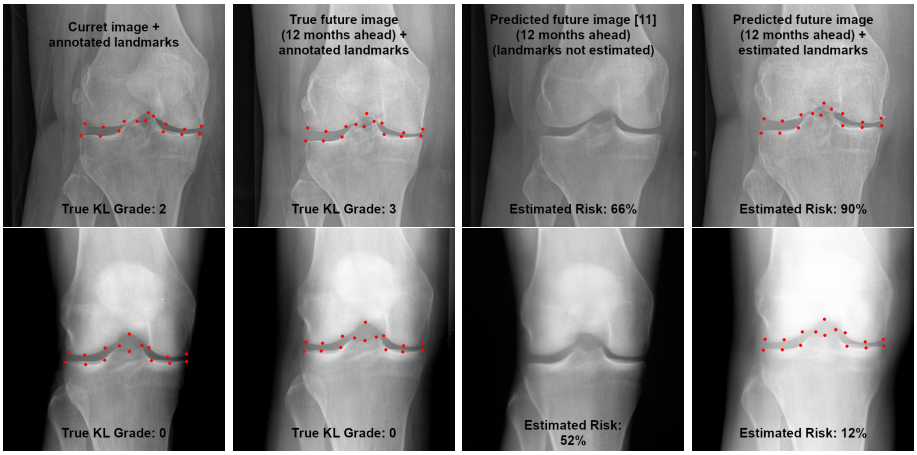

Fig. 3. (Left) Current image with annotated landmarks. (Centre Left) True future image (12 months ahead) with annotated landmarks. (Centre Right) Predicted future image by [9] (note that it does not show landmarks) (Right) Predicted future image (12 months ahead) with estimated landmarks by our method. (Top) Progressing OA with KL grade from 2 to 3. (Bottom) No OA with KL grade 0.

5 Discussion & Conclusion

The proposed method achieves $\sim 9\times$ faster inference and a higher risk estimation AUC (0.71 vs. 0.69) than the current SOTA [9]. By utilising a class-conditioned latent space, our approach enables diffusion models to generate images suitable for predicting future disease progression and allows for a more compact model than the SOTA (our model has 35M vs. 215M parameters in [9]), and especially in comparison with similar methods utilising diffusion models (35M vs. 1.1B in [30]). Furthermore, incorporating anatomical knee landmarks improves risk estimation while providing additional interpretable outputs.

We find that test-time upscaling of $\mathbf{z}^0$ improves risk estimation for stable low KL scores (0→0,1→1) and increasing high scores (2→3, 3→4) but worsens stable high scores (3→3,4→4) and increasing low scores (0→1, 1→2). We hypothesise this stems from resizing-induced bias, as joint spacing depends on size, whereas osteophytes and sclerosis are more influenced by texture and opacity. Additionally, the model struggles with KL class 1, likely due to its inherent ambiguity—representing doubtful cases rather than mild osteoarthritis—introducing noise that affects neighboring classes (0 and 2). In contrast, clearer symptoms classes (3 and 4) achieve the highest classification accuracy, emphasizing the need for improved label noise handling.

The main limitation of our method is its dependence on class and landmark annotations, which may not always be available. However, landmark annotations are only useful for risk estimation, not for image generation.

For future work, our flexible conditioning mechanism could be extended to multi-image inputs, such as both knees or prior exams, to improve progression modelling. Additionally, recent advancements in conditioning latent diffusion models with non-image data could be explored to enhance predictions. Finally, iterative risk estimation could allow for longer-term forecasting beyond 12 months, improving clinical applicability.

Acknowledgments. This work was partly supported by the Engineering and Physical Sciences Research Council (EPSRC) through grant EP/Y018036/1.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alaluf, Y., et al.: Only a matter of style: Age transformation using a style-based regression model. ACM Transactions on Graphics **40** (7 2021). https://doi.org/10.1145/3450626.3459805
2. Bhagwat, N., et al.: Modeling and prediction of clinical symptom trajectories in alzheimer's disease using longitudinal data. PLOS Computational Biology **14**(9), 1–25 (09 2018). https://doi.org/10.1371/journal.pcbi.1006376, https://doi.org/10.1371/journal.pcbi.1006376
3. Chu, C.R., et al.: Early diagnosis to enable early treatment of pre-osteoarthritis. Arthritis Research & Therapy **14**, 212 (2012). https://doi.org/10.1186/ar3845, https://doi.org/10.1186/ar3845
4. Cigdem, O., et al.: Estimation of time-to-total knee replacement surgery (2024), https://arxiv.org/abs/2405.00069
5. Eckstein, F., Wirth, W., Nevitt, M.C.: Recent advances in osteoarthritis imaging–the osteoarthritis initiative. Nat. Rev. Rheumatol. **8**(10), 622–630 (Oct 2012)
6. Esser, P., Rombach, R., Ommer, B.: Taming transformers for high-resolution image synthesis. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 12868–12878. IEEE Computer Society (6 2021). https://doi.org/10.1109/CVPR46437.2021.01268, https://doi.ieeecomputersociety.org/10.1109/CVPR46437.2021.01268
7. Gafuroglu, C., Rekik, I.: Image evolution trajectory prediction and classification from baseline using learning-based patch atlas selection for early diagnosis. CoRR **abs/1907.06064** (2019), http://arxiv.org/abs/1907.06064
8. Gogna, A., Majumdar, A., Ward, R.K.: Semi-supervised stacked label consistent autoencoder for reconstruction and analysis of biomedical signals. IEEE Transactions on Biomedical Engineering **64**, 2196–2205 (2017), https://api.semanticscholar.org/CorpusID:3543838
9. Han, T., et al.: Image prediction of disease progression for osteoarthritis by style-based manifold extrapolation. Nature Machine Intelligence **4**(11), 1029–1039 (Nov 2022). https://doi.org/10.1038/s42256-022-00560-x, https://doi.org/10.1038/s42256-022-00560-x
10. He, R., Ang, G., Tward, D.: Individualized multi-horizon mri trajectory prediction for alzheimer's disease (9 2024)

11. Huang, Y., et al.: Longitudinal prediction of postnatal brain magnetic resonance images via a metamorphic generative adversarial network. Pattern Recognition **143** (11 2023). https://doi.org/10.1016/j.patcog.2023.109715
12. Jin, C., et al.: Predicting treatment response from longitudinal images using multi-task deep learning. Nature Communications **12** (12 2021). https://doi.org/10.1038/s41467-021-22188-y
13. Kohn, M.D., et al.: Classifications in brief: Kellgren-lawrence classification of osteoarthritis. Clin. Orthop. Relat. Res. **474**, 1886–1893 (8 2016)
14. Lauritzen, A.D., et al.: Assessing breast cancer risk by combining ai for lesion detection and mammographic texture. Radiology **308**, e230227 (8 2023). https://doi.org/10.1148/radiol.230227
15. Liu, C., et al.: Imageflownet: Forecasting multiscale trajectories of disease progression with irregularly-sampled longitudinal medical images (9 2024). https://doi.org/10.36227/techrxiv.172297920.01199828/v1
16. Lu, X.H., et al.: Recurrent disease progression networks for modelling risk trajectory of heart failure. PLOS ONE **16**(1), 1–15 (01 2021). https://doi.org/10.1371/journal.pone.0245177, https://doi.org/10.1371/journal.pone.0245177
17. Nguyen, H.H., et al.: Clinically-inspired multi-agent transformers for disease trajectory forecasting from multimodal data. IEEE Transactions on Medical Imaging **43**, 529–541 (2024). https://doi.org/10.1109/TMI.2023.3312524
18. Nguyen, K.P., et al.: Predicting parkinson's disease trajectory using clinical and neuroimaging baseline measures. Parkinsonism & Related Disorders **85**, 44–51 (2021). https://doi.org/https://doi.org/10.1016/j.parkreldis.2021.02.026, https://www.sciencedirect.com/science/article/pii/S1353802021000754
19. Or-El, R., et al.: Lifespan age transformation synthesis. In: Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI. pp. 739–755. Springer-Verlag (2020). https://doi.org/10.1007/978-3-030-58539-6_44, https://doi.org/10.1007/978-3-030-58539-6$_4$4
20. Placido, D., et al.: A deep learning algorithm to predict risk of pancreatic cancer from disease trajectories. Nature Medicine **29**, 1113–1122 (2023). https://doi.org/10.1038/s41591-023-02332-5, https://doi.org/10.1038/s41591-023-02332-5
21. Rasheed, K., et al.: Explainable, trustworthy, and ethical machine learning for healthcare: A survey. Computers in Biology and Medicine **149**, 106043 (2022). https://doi.org/https://doi.org/10.1016/j.compbiomed.2022.106043, https://www.sciencedirect.com/science/article/pii/S0010482522007569
22. Rombach, R., et al.: High-resolution image synthesis with latent diffusion models. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10674–10685. IEEE Computer Society (6 2022). https://doi.org/10.1109/CVPR52688.2022.01042, https://doi.ieeecomputersociety.org/10.1109/CVPR52688.2022.01042
23. Salimans, T., Ho, J.: Progressive distillation for fast sampling of diffusion models. In: International Conference on Learning Representations (2022), https://openreview.net/forum?id=TIdIXIpzhoI
24. Shibata, H., et al.: Aging prediction using deep generative model toward the development of preventive medicine (9 2022). https://doi.org/10.48550/arXiv.2208.10797
25. Sinusas, K.: Osteoarthritis: diagnosis and treatment. Am. Fam. Physician **85**, 49–56 (1 2012)

26. Tariq, A., et al.: Graph-based fusion modeling and explanation for disease trajectory prediction (9 2022). https://doi.org/10.1101/2022.10.25.22281469
27. Tariq, A., et al.: Fusion of imaging and non-imaging data for disease trajectory prediction for coronavirus disease 2019 patients. Journal of Medical Imaging **10**, 34004 (2023). https://doi.org/10.1117/1.JMI.10.3.034004, https://doi.org/10.1117/1.JMI.10.3.034004
28. Tariq, A., et al.: Generalizable model design for clinical event prediction using graph neural networks. medRxiv (2023). https://doi.org/10.1101/2023.03.22.23287599, https://www.medrxiv.org/content/early/2023/03/25/2023.03.22.23287599
29. Tiulpin, A., et al.: Kneel: Knee anatomical landmark localization using hourglass networks. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). pp. 352–361 (2019). https://doi.org/10.1109/ICCVW.2019.00046
30. Zhao, Z., et al.: Extrapolating Prospective Glaucoma Fundus Images through Diffusion in Irregular Longitudinal Sequences . In: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 4032–4035. IEEE Computer Society, Los Alamitos, CA, USA (Dec 2024). https://doi.org/10.1109/BIBM62325.2024.10822368, https://doi.ieeecomputersociety.org/10.1109/BIBM62325.2024.10822368