Foundational Resources for the Definitive Master Blueprint: A Technical and Scientific Compendium

Introduction

Objective

This report provides a curated and contextualized set of foundational resources for each component of the RSNA 2025 Intracranial Aneurysm Detection challenge blueprint. The aim is to go beyond a simple bibliography, offering strategic analysis and actionable intelligence by synthesizing seminal papers, high-quality codebases, and comprehensive reviews. The analysis is informed by the strategic rationale outlined in the project's internal planning document, "A Physics-Informed, Oscillatory Network Strategy for Aneurysm Detection".[1] Each section is designed to equip the development team with the necessary scientific background, practical implementation details, and broader contextual understanding to execute the blueprint effectively.

Methodology

The report strictly follows the structure of the "Master Blueprint." For each numbered component, it presents the primary academic papers that introduced the core concept, a direct link to a robust and well-regarded open-source implementation (with a preference for PyTorch), and references to recent survey literature. This structure ensures that the resources are directly mapped to the project's technical requirements, facilitating a seamless transition from research to development.

Part I: Foundational Asset Creation

This section details the resources required for the critical preparatory steps of the project: pre-training the core feature extractors and establishing the foundational models for auxiliary tasks such as vessel segmentation. The quality of these assets is paramount, as they form the bedrock upon which the entire detection pipeline is built.

1.1 Pre-trained Backbone: SparK & Hierarchical Masking (MiM)

The chosen pre-training strategy combines a state-of-the-art self-supervised learning (SSL) algorithm, SparK, with an advanced, anatomy-aware masking strategy, MiM. This pairing is designed to create a powerful feature extractor that is not only robust to data scarcity but also possesses a deep, multi-scale understanding of vascular anatomy.

1.1.1 SparK: Sparse Masked Modeling on Images via Sparse Convolutions

- **Primary Foundational Paper(s)**
  - **Paper:** Tian, K., Jiang, Y., Diao, Q., Lin, C., Wang, L., & Yuan, Z. (2023). "Designing BERT for Convolutional Networks: Sparse and Hierarchical Masked Modeling". *arXiv:2301.03580*.
  - **Significance:** This paper introduces **SparK (Sparse masKed modeling)**, a generative SSL method that successfully adapts the masked image modeling (MIM) paradigm, popularized by Vision Transformers (ViTs), to convolutional neural networks (ConvNets). It overcomes two key challenges that previously hindered this adaptation: (1) the inability of standard dense convolutions to process irregularly masked inputs efficiently, and (2) the single-scale nature of the original BERT-style pre-training, which is inconsistent with the hierarchical structure of ConvNets. The core innovation is to treat the unmasked pixels as a sparse set of voxels and process them using highly efficient sparse convolutions.[2] This approach is general and can be applied to any standard ConvNet architecture, such as ResNet or ConvNeXt, without requiring modifications to the backbone. The paper demonstrates that this method surpasses both state-of-the-art contrastive learning and Transformer-based MIM on numerous downstream tasks.[2]
  - **Strategic Rationale:** The selection of SparK is underpinned by a coherent philosophy of building a robust and resilient model. Internal project analysis correctly identifies SparK's "superior robustness in data-scarce fine-tuning scenarios" as its most critical advantage.[1] Empirical studies show that SparK's performance remains stable even when the downstream training data is reduced by 50%, whereas the performance of leading contrastive methods degrades significantly under the same conditions.[1] This resilience is a crucial de-risking factor for a competition that involves an unseen private test set, where rare aneurysm subtypes may constitute a data-scarce environment. SparK's focus on reconstructing fine-grained local structures provides a more resilient foundation than methods that rely on instance-level discrimination.

- **High-Quality Implementations**
  - **Official Repository (PyTorch):** https://github.com/keyu-tian/SparK
  - **Analysis:** This is the official implementation from the paper's authors, providing the complete source code, training scripts, and pre-trained models for both classical (ResNet) and modern (ConvNeXt) architectures.[2] The repository is well-documented and is designed to integrate with the popular mmpretrain library, which offers clear and standardized procedures for training and inference. This will facilitate its seamless integration into the project's development pipeline.[2]
- **Recent Survey/Review Papers**
  - **Paper 1:** Hondru, V., & Ionescu, R. T. (2024). "Masked Image Modeling: A Survey". *International Journal of Computer Vision*.
  - **Paper 2:** Xie, Z., Geng, Z., Hu, J., et al. (2023). "Revealing the Dark Secrets of Masked Image Modeling". In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
  - **Significance:** While no surveys focus exclusively on the SparK algorithm, these papers provide excellent context on the broader field of Masked Image Modeling (MIM), of which SparK is a key convolutional example. They categorize MIM approaches (e.g., reconstruction-based vs. contrastive), detail the common encoder-decoder framework, and discuss the evolution of the paradigm from its origins in Natural Language Processing (BERT) to its application in computer vision.[4] The work by Xie et al. [4] offers particularly deep analysis into
    why MIM is so effective, finding that it imparts a strong "locality inductive bias" to all layers of a model. This forces the model to learn rich local and contextual information, which is particularly beneficial for Vision Transformers and aligns perfectly with the project's goal of learning the fine-grained anatomical features necessary for aneurysm detection.[1]

1.1.2 Hierarchical Masking (MiM) for 3D Anatomical Context

- **Primary Foundational Paper(s)**
  - **Paper:** Zhuang, J., Wu, L., Wang, Q., Fei, P., Vardhanabhuti, V., Luo, L., & Chen, H. (2024). "MiM: Mask in Mask Self-Supervised Pre-Training for 3D Medical Image Analysis". *arXiv:2404.15580*.
  - **Significance:** This paper introduces the **Mask in Mask (MiM)** framework, a hierarchical masking strategy designed specifically for the challenges of 3D

medical imaging.[7] It advances standard Masked Autoencoder (MAE) approaches by explicitly modeling the intrinsically hierarchical nature of anatomical structures. The framework creates multi-level masked inputs; for example, a coarse mask is applied at a global scale (Level 1), and then a second, finer-grained mask is applied *within* the previously masked regions to create a local-scale input (Level 2). The model is tasked with reconstructing the masked tokens at all levels simultaneously. Furthermore, a cross-level alignment loss, typically a contrastive loss, is used to enforce semantic consistency between features learned at adjacent scales.[1]
    - **Strategic Rationale:** The blueprint's choice of MiM is strategically sound and demonstrates a nuanced understanding of the problem domain. As noted in the internal project document, aneurysm detection is an "intrinsically multi-scale problem" that requires the identification of a small, local morphological anomaly (the aneurysm) and the simultaneous understanding of its global topological context within the cerebrovascular tree.[1] MiM's nested masking scheme and cross-level alignment provide a powerful inductive bias for learning the tree-like, hierarchical structure of the vasculature. This pre-training objective is perfectly aligned with the downstream task, giving it a significant advantage over single-scale or anatomy-agnostic masking strategies.[1]
- **High-Quality Implementations**
    - **Related Repository (PyTorch):** https://github.com/ge-xing/HybridMIM
    - **Analysis:** The official "Mask in Mask" paper states that the code will be made available upon the paper's acceptance.[7] As of the writing of this report, a dedicated official repository has not been identified. However, the HybridMIM repository implements a closely related concept of hierarchical and hybrid masked modeling for 3D medical image segmentation.[9] It supports both UNet and SwinUNETR backbones and focuses on learning from multiple levels (pixel, region, and sample). While not a direct implementation of the Zhuang et al. paper, it represents a high-quality, relevant starting point for implementing the hierarchical masking concept in PyTorch. Other curated lists of MIM resources, such as "Awesome-MIM," provide extensive links but do not yet contain an official implementation for this specific paper.[10]
- **Recent Survey/Review Papers**
    - **Paper 1:** Chen, J., et al. (2023). "Masked Image Modeling Advances 3D Medical Image Analysis". In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*.
    - **Paper 2:** Tang, Y., et al. (2024). "GMIM: Self-supervised pre-training for 3D

medical image segmentation with adaptive and hierarchical masked image modeling". *Computers in Biology and Medicine*.
    - **Significance:** These reviews focus specifically on the application of MIM within the 3D medical imaging domain. The work by Chen et al. [12] was one of the first to systematically study MIM for 3D medical tasks, demonstrating that it leads to significant speed-ups in training convergence and ultimately improves downstream segmentation performance. The GMIM paper [13] introduces another, similar hierarchical masking strategy, reinforcing the trend and validity of the MiM approach for learning the correlations between different anatomical structures. These papers collectively affirm that moving beyond simple random masking toward more intelligent, anatomy-aware hierarchical strategies is crucial for unlocking the full potential of self-supervised learning on 3D medical volumes.[12]

1.2 Pre-trained Backbone: WaveFormer

- **Primary Foundational Paper(s)**
    - **Paper:** Perera, S., Navard, P., & Yilmaz, A. (2024). "WaveFormer: A 3D Transformer with Wavelet-Driven Feature Representation for Efficient Medical Image Segmentation". In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*.
    - **Significance:** This paper introduces WaveFormer, a novel 3D transformer architecture specifically designed to be computationally efficient for 3D medical image segmentation.[3] The core innovation is the integration of the Discrete Wavelet Transform (DWT) into the transformer block. The DWT is used to decompose 3D feature maps into low-frequency (approximations, representing global context) and high-frequency (details, representing fine-grained local features) sub-bands. The computationally expensive multi-head self-attention mechanism is then applied *only* to the compact, low-frequency component, which drastically reduces the computational and memory load. The decoder subsequently uses the Inverse DWT (IDWT) for efficient, high-fidelity upsampling.[3] This design is biologically inspired by the top-down mechanisms of the human visual system and is shown to achieve state-of-the-art performance with significantly fewer parameters than competing models.[15]
- **High-Quality Implementations**
    - **Identified Gap:** A significant finding of this report is a discrepancy between

the project blueprint's requirements and publicly available code. While the blueprint specifies the 3D medical imaging WaveFormer, the most prominent public GitHub repository with the name "Waveformer" (https://github.com/vb000/Waveformer) is an implementation for an unrelated *low-latency audio processing* task.[17] The foundational paper for the medical WaveFormer does not provide a direct link to an official repository, and extensive searches of other 3D medical segmentation repositories have not yielded a public implementation.[16]
    - **Recommendation:** This represents a critical implementation gap and a potential project risk. The development team must be prepared to implement the WaveFormer architecture directly from the detailed specifications provided in the Perera et al. (2024) paper. The paper provides a clear architectural breakdown of the encoder and decoder stages, including the wavelet-attention blocks, which will be essential for this task.[3]
- **Recent Survey/Review Papers**
    - **Paper 1:** He, K., et al. (2024). "A review of deep learning in 3D semantic segmentation for 3D point clouds". *Remote Sensing*.
    - **Paper 2:** Various Authors. (2024). "Deep learning in medical image segmentation: a review". *Informatics in Medicine Unlocked*.
    - **Significance:** These reviews comprehensively cover the landscape of 3D semantic segmentation, particularly within medical imaging.[21] They chronicle the evolution from purely CNN-based models (like U-Net and its variants) to the adoption of Transformer-based architectures. A central theme in this literature is the trade-off between the limited receptive field of CNNs and the quadratic computational cost of Vision Transformers, especially when applied to high-resolution 3D volumes.[22] The WaveFormer architecture fits squarely into this narrative as an "efficient Transformer" that attempts to capture long-range dependencies without incurring prohibitive computational costs. These surveys provide the broader context of the key challenges in 3D medical segmentation—such as high memory footprints, the need for computational efficiency, and the difficulty of capturing both local and global features—which WaveFormer is explicitly designed to address.[23]

1.3 Vessel Segmentation: Tversky Loss

- **Primary Foundational Paper(s)**
    - **Paper:** Salehi, S. S. M., Erdogmus, D., & Gholipour, A. (2017). "Tversky loss

function for image segmentation using 3D fully convolutional deep networks". In *International Workshop on Machine Learning in Medical Imaging*.
  - **Significance:** This paper introduces a loss function based on the Tversky Index, specifically designed to address the problem of severe class imbalance in medical image segmentation.[25] The Tversky Index is a generalization of the widely used Dice coefficient and Tanimoto coefficient. Its key feature is the introduction of two tunable parameters,
    α and β, which allow for the differential weighting of false positives (FPs) and false negatives (FNs) in the loss calculation.[25] The Tversky Index (
    TI) is given by the formula:

    TI(α,β)=TP+αFN+βFPTP

    where TP is the number of true positives. By adjusting α and β, one can control the trade-off between precision and recall during training.
  - **Strategic Rationale:** The project's internal analysis correctly identifies the highly asymmetric sensitivity of the two-stage pipeline to segmentation errors. A false negative (a missed vessel) is described as a "catastrophic failure," as it prevents the downstream classifier from ever seeing the potential aneurysm. Conversely, a false positive (over-segmentation) is "suboptimal but non-catastrophic".[1] The Tversky loss directly addresses this strategic priority. By setting
    β>α (e.g., β=0.7,α=0.3 as recommended in the blueprint), the loss function applies a heavier penalty to false negatives (missed vessel voxels). This explicitly optimizes the auxiliary segmentation model to maximize recall (sensitivity), thereby mitigating the most critical failure mode for the entire pipeline.[1]
- **High-Quality Implementations**
  - **PyTorch (Kaggle Kernel):**
    https://www.kaggle.com/code/bigironsphere/loss-function-library-keras-pytorch
  - **Analysis:** This popular and well-maintained Kaggle kernel provides a clean, backend-native PyTorch implementation of the Tversky Loss.[27] It is presented as a Python class,
    TverskyLoss(nn.Module), that is easy to import and integrate into any standard PyTorch training loop. The code is clear, well-commented, and directly implements the formula from the foundational paper. The author also provides valuable practical advice based on extensive experimentation, such as noting that Tversky and related loss functions often benefit from very low

learning rates (e.g., 5e−5 to 1e−4) and may require more epochs to show significant improvement.[27]

- **Recent Survey/Review Papers**
  - **Paper 1:** Ma, J. (2021). "Loss odyssey in medical image segmentation". *Medical Image Analysis*.
  - **Paper 2:** Taha, B., & Al-antari, M. A. (2025). "A Comprehensive Review of Loss Functions in Medical Image Segmentation". *Journal of Imaging*.
  - **Significance:** These reviews provide a comprehensive taxonomy of the numerous loss functions used in medical image segmentation. They categorize losses based on their derivation and primary objective, such as distribution-based (e.g., Cross-Entropy), region-based (e.g., Dice, IoU, Tversky), boundary-based, and compound losses.[28] They place Tversky Loss within the broader context of region-based losses, highlighting its specific utility as a generalization of Dice Loss for handling significant class imbalance.[28] The empirical study by Ma (2021) is particularly valuable, as it finds that compound loss functions (e.g., a combination of Dice with another loss like Focal Loss or Boundary Loss) are often the most robust across different tasks. This suggests a potential avenue for future enhancement of the blueprint's segmentation model.[28]

Part II: Candidate Generation ("The Smart Magnetic Net")

This phase of the blueprint focuses on the intelligent generation of high-quality aneurysm candidates. It moves beyond simple intensity-based blob detection to a more sophisticated, topologically-aware analysis of the cerebrovasculature, leveraging a combination of an efficient 3D segmentation model and a graph-based analytical framework.

2.1 Morphological Extraction: SegFormer3D

- **Primary Foundational Paper(s)**
  - **Paper:** Perera, S., Navard, P., & Yilmaz, A. (2024). "SegFormer3D: An Efficient Transformer for 3D Medical Image Segmentation". In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*

(CVPRW).
    - **Significance:** This paper introduces SegFormer3D, a hierarchical Transformer architecture specifically designed for efficient 3D medical image segmentation.[31] It extends the principles of the successful 2D SegFormer by using a lightweight, all-MLP decoder to aggregate multi-scale features from a Transformer encoder. This design avoids the complex and parameter-heavy decoders common in other architectures. The result is a model that is significantly more efficient in terms of both parameters and computation (e.g., offering a 33x reduction in parameters and a 13x reduction in GFLOPS compared to some state-of-the-art models) while maintaining highly competitive performance on standard benchmarks like BraTS, Synapse, and ACDC.[31]
    - **Strategic Rationale:** For the candidate generation stage of the pipeline, SegFormer3D's computational efficiency is a major strategic asset. This stage must process full 3D volumes quickly and accurately to produce a vessel probability map that will serve as the input for the subsequent topological analysis. A lightweight yet powerful model like SegFormer3D is an ideal choice to perform this initial morphological extraction without consuming an excessive amount of the computational budget, which is better reserved for the more complex Expert Classifier in the next stage.
- **High-Quality Implementations**
    - **Official Repository (PyTorch):** https://github.com/OSUPCVLab/SegFormer3D
    - **Analysis:** This is the official PyTorch implementation from the authors of the CVPRW 2024 paper.[32] The repository is comprehensive, providing the model architecture (
      architectures/segformer3d.py), configuration files for reproducing the paper's results, and detailed instructions for training and evaluation on standard medical datasets.[34] Notably, the repository includes pre-processing scripts that were adapted from the influential nnFormer repository. This was done to ensure a fair and direct comparison with baseline architectures, which adds to the credibility and practical utility of the provided codebase.[34]
- **Recent Survey/Review Papers**
    - Refer to Section 1.2 for relevant survey papers. The same reviews on efficient transformers and 3D medical segmentation provide the necessary context for understanding why lightweight yet powerful models like SegFormer3D represent an important and active area of research.

2.2 Topological Analysis: Vessel Graph Networks (VGNs)

This sub-section covers the set of components needed to execute the core innovation of the "Smart Magnetic Net": converting the dense, voxel-based segmentation into a sparse graph representation of the vasculature and analyzing its topology for anomalies.

2.2.1 Vessel Graph Network (VGN) Framework

- **Primary Foundational Paper(s)**
  - **Paper:** Shin, S. Y., Lee, S., Yun, I. D., & Lee, K. M. (2019). "Deep vessel segmentation by learning graphical connectivity". *Medical Image Analysis*.
  - **Significance:** This paper proposes a unified framework that incorporates a Graph Neural Network (GNN) into a standard CNN architecture to improve vessel segmentation.[35] The key idea is to move beyond purely local, pixel-based predictions and to explicitly leverage the global, graph-like structure of the vessel network. The GNN component acts as a regularizer, learning the relationships between vessel neighborhoods to enforce connectivity, resolve ambiguities in low-contrast regions, and reduce fragmented predictions. This results in more topologically correct segmentations, especially for fine or broken vessels.[35]
  - **Strategic Rationale:** The "Smart Magnetic Net" directly builds upon the philosophy pioneered in this paper. The blueprint uses a primary segmentation model (SegFormer3D) to extract the initial vessel structure, and then explicitly models it as a graph to find *topological anomalies* (aneurysms), rather than just intensity-based anomalies. This paper provides the foundational concept for this powerful hybrid CNN-GNN approach to vascular analysis.
- **High-Quality Implementations**
  - **Official Repository (PyTorch):** https://github.com/syshin1014/VGN
  - **Analysis:** This is the official PyTorch implementation for the "Deep Vessel Segmentation by Learning Graphical Connectivity" paper.[37] The repository includes code for the full pipeline described in the paper: pre-training the

CNN backbone, constructing the graphs from the resulting probability maps, and training the full VGN model. Although the original work focused on 2D retinal images, the underlying principles and code structure for graph construction and GNN training are directly applicable to the 3D cerebrovascular use case and provide a strong foundational codebase.

2.2.2 3D Morphological Skeletonization

- **Primary Algorithm(s) and Implementations**
  - **Algorithm:** 3D Morphological Thinning (e.g., Zhang-Suen or Lee's algorithm).
  - **High-Quality Implementation (Python):** skimage.morphology.skeletonize_3d
  - **Significance:** Skeletonization is the critical algorithmic step that converts the 3D binary vessel mask into a 1-voxel-thick centerline representation. This process, also known as thinning, preserves the essential connectivity and topology of the vessel tree and forms the direct basis for constructing the graph nodes (junctions and endpoints) and edges (connecting paths).[1] The scikit-image library, a staple of the scientific Python ecosystem, provides a robust, well-tested, and efficient implementation of 3D skeletonization.[38] The library's documentation notes that for 3D images, it automatically selects Lee's (1994) method, which is specifically designed for 3D volumes. This method uses an octree data structure to examine a $3 \times 3 \times 3$ neighborhood and employs an iterative process to remove border pixels without breaking the object's connectivity, making it ideal for this task.[38]

2.2.3 Graph Attention Networks (GATs)

- **Primary Foundational Paper(s)**
  - **Paper:** Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2018). "Graph Attention Networks". In *International Conference on Learning Representations (ICLR)*.
  - **Significance:** This seminal paper introduced Graph Attention Networks (GATs), a novel neural network architecture that operates on graph-structured data by leveraging masked self-attentional layers.[39] This was a significant

departure from prior methods like Graph Convolutional Networks (GCNs), which use fixed, structurally-defined weights (e.g., based on node degree) for aggregating information from neighboring nodes. In contrast, GATs learn to assign different importance weights to different nodes within a neighborhood during the aggregation process.[1] This allows the model to dynamically focus on the most relevant parts of the neighborhood for a given task, without requiring costly matrix operations or depending on knowing the graph structure upfront.
    - **Strategic Rationale:** The internal project document makes a compelling case for using GATs for the specific task of topological anomaly detection.[1] An aneurysm represents a localized anomaly—a sudden, abnormal expansion in the vessel graph. A standard GCN, which performs an isotropic aggregation of neighbor features, might "smooth over" or dilute the features of this anomaly by averaging them with its normal neighbors. A GAT, however, can learn the complex, context-dependent rules of healthy vascular topology. It can learn to assign a very high attention weight to an anomalous neighboring node (e.g., one with an unusually large radius connected by a very short edge), causing the resulting node embedding to be significantly different from that of a healthy bifurcation. This makes the anomalous node easily separable in the feature space, making GATs an exceptionally well-suited architecture for this task.[1]
- **High-Quality Implementations**
    - **PyTorch (Annotated Implementation):** https://nn.labml.ai/graphs/gat/index.html
    - **Analysis:** This resource provides a clear, extensively annotated PyTorch implementation of a GraphAttentionLayer built from scratch, closely following the methodology of the original paper.[42] The code breaks down the forward pass into logical, step-by-step components, explaining the initial linear transformations, the concatenation of node features for attention score calculation, the application of the attention mask based on the adjacency matrix, and the final softmax normalization. This serves as an excellent educational resource for understanding the mechanics of GATs and provides a solid, well-explained foundation for building a custom GAT model for the project.

2.2.4 Contextual Surveys on Graph Neural Networks in Medical Imaging

- **Paper 1:** Kazi, A., et al. (2021). "A Systematic Review of Graph Neural Networks in Medical Imaging". *Frontiers in Artificial Intelligence*.
- **Paper 2:** Shehata, M., et al. (2022). "A Critical Review of the State-of-the-art in Medical Shape Analysis using GNNs". In *MICCAI Workshop on Shape in Medical Imaging*.
- **Significance:** These reviews situate the use of GNNs within the broader medical imaging context, providing a strong justification for their inclusion in the blueprint.[43] They highlight that GNNs are a natural fit for modeling the non-Euclidean data that is common in medicine, such as anatomical structures (like the vasculature), functional brain connectivity, and molecular interactions.[44] They discuss the paradigm shift from CNNs, which are fundamentally limited to grid-like data, to GNNs that can capture complex, irregular relationships between entities. The review by Kazi et al. provides a systematic overview of GNN applications, while the work by Shehata et al. conducts a valuable comparative analysis of different GNN architectural choices (e.g., different convolutional layers) for shape classification, offering practical insights for model design.[43]

Part III: Candidate Classification ("The Expert Classifier")

This phase of the blueprint details the resources for the sophisticated hybrid classifier responsible for the final, high-precision analysis of the candidate regions generated by the Smart Magnetic Net. This stage employs a novel architecture that integrates principles from dynamical systems and fluid dynamics to move beyond simple pattern recognition.

3.1 Dynamic Analysis: Artificial Kuramoto Oscillatory Neurons (AKOrN)

- **Primary Foundational Paper(s)**
  - **Paper:** Miyato, T., et al. (2025). "Artificial Kuramoto Oscillatory Neurons". In *International Conference on Learning Representations (ICLR)*.
  - **Significance:** This paper introduces **Artificial Kuramoto Oscillatory Neurons (AKOrN)**, a novel and dynamic alternative to standard static activation units like ReLU or GeLU.[46] The architecture is based on the Kuramoto model, a mathematical model from physics that describes

synchronization phenomena in systems of coupled oscillators. Instead of applying a simple threshold function, AKOrN layers consist of dynamic, interacting oscillators that evolve on the surface of a hypersphere. The key computational mechanism is synchronization, which allows the network to dynamically "bind" related features together into a single, coherent representation.[1] The paper demonstrates that this approach leads to significant performance improvements on a wide range of tasks that require reasoning, unsupervised object discovery, and adversarial robustness.[47]

  - **Strategic Rationale:** The use of AKOrN represents a significant conceptual advance over standard deep learning architectures. The blueprint proposes placing the AKOrN block after the transformer's MLP block, allowing the oscillatory dynamics to operate on highly processed, context-aware feature vectors. The goal is to use the synchronization dynamics to group and refine these abstract features—for example, binding the features corresponding to a vessel wall with the features corresponding to turbulent flow within it to form a single, unified representation of an "aneurysm".[1] This use of a biologically-inspired dynamical system as a computational primitive injects a powerful inductive prior into the model, encouraging it to learn more abstract and robust representations.
- **High-Quality Implementations**
  - **Official Repository (PyTorch):** https://github.com/autonomousvision/akorn
  - **Analysis:** This is the official PyTorch implementation from the authors of the ICLR 2025 paper, provided by the Autonomous Vision Group at the University of Tübingen and MPI-IS.[48] The repository provides the core source code for the AKOrN layers and includes examples of their integration into various network architectures. This is the definitive and essential resource for implementing this component. It is important to note that other public repositories with the name "acorn" exist but are for unrelated projects, such as a programming language or an LLM agent framework, and should be disregarded.[50]
- **Recent Survey/Review Papers**
  - **Context:** AKOrN is a very recent and novel concept, introduced in late 2024 for the ICLR 2025 conference. As such, dedicated survey papers reviewing this specific architecture do not yet exist. The foundational paper itself provides the best contextualization, relating the work to modern dynamical views of neurons, neuroscience concepts like traveling waves and feature binding, and connections to physics-based energy models.[48]

3.2 Contextual Fusion: Gated Attention (FiLM-style)

- **Primary Foundational Paper(s)**
  - **Paper:** Perez, E., Strub, F., de Vries, H., Dumoulin, V., & Courville, A. (2018). "FiLM: Visual Reasoning with a General Conditioning Layer". In *AAAI Conference on Artificial Intelligence*.
  - **Significance:** This paper introduces **FiLM (Feature-wise Linear Modulation)**, a simple yet powerful general-purpose method for conditioning neural networks.[52] A FiLM layer performs a feature-wise affine transformation ($y=\gamma\cdot x+\beta$) on the intermediate feature maps of a network. The key insight is that the scaling parameters (γ) and shifting parameters (β) are not learned directly but are instead generated by a separate, smaller network that takes some conditioning information as input (e.g., a text query, or in this project's case, patient metadata).[52] This allows the conditioning information to dynamically modulate the computation of the main network at a deep feature level, effectively acting as a "gate" or controller for feature processing.
  - **Strategic Rationale:** The blueprint proposes using a FiLM-style gated attention mechanism to fuse patient metadata (e.g., age, sex) with the image features extracted by the backbone.[1] This approach is far superior to the common but suboptimal strategy of simple late-fusion (i.e., concatenating the metadata vector with the flattened image features just before the final classification layer). FiLM allows the model to learn complex, non-linear interactions between the patient's characteristics and the visual evidence. It effectively enables the metadata to re-weigh or gate entire feature channels, allowing the model to learn, for instance, that certain textural features are more indicative of an aneurysm in older patients and should therefore be amplified when the "age" metadata value is high.[1]
- **High-Quality Implementations**
  - **PyTorch Geometric:** torch_geometric.nn.conv.FiLMConv
  - **Analysis:** While the original FiLM paper focused on vision-language tasks, the mechanism itself is highly general. A high-quality, well-tested implementation exists within the popular PyTorch Geometric library as FiLMConv.[55] This class implements the FiLM mechanism in the context of graph neural networks, but the core logic of generating β and γ parameters to modulate features is directly transferable to the 3D

convolutional or transformer backbone used in this project. It provides a robust and reliable starting point for implementing this fusion strategy. The continued relevance of FiLM is demonstrated by its successful application in more recent GNN architectures as well.[56]

- **Recent Survey/Review Papers**
  - **Paper:** Ramirez, J., et al. (2020). "Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines". *Journal of the American Medical Informatics Association*.
  - **Significance:** This systematic review [53] specifically covers the critical challenge of fusing imaging data with non-imaging data like Electronic Health Records (EHRs). It provides a valuable taxonomy of fusion strategies, categorizing them as early (input-level), late (decision-level), or intermediate fusion. The review highlights methods like gated attention and FiLM-style modulation as advanced intermediate fusion techniques that provide a much richer way for a model to integrate heterogeneous information compared to simple concatenation. This directly validates the advanced approach chosen in the project blueprint.

3.3 Physics Regularizer: Physics-Informed Neural Networks (PINNs)

The integration of a PINN as an auxiliary head is a core innovation of the Expert Classifier, designed to regularize the model by ensuring its latent representations are consistent with the physical laws of fluid dynamics.

3.3.1 Foundational PINN Framework (Raissi et al.)

- **Primary Foundational Paper(s)**
  - **Paper:** Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations". *Journal of Computational Physics*.
  - **Significance:** This is the seminal paper that introduced and popularized the modern **Physics-Informed Neural Network (PINN)** framework.[57] It established the core idea of embedding a partial differential equation (PDE)

directly into the loss function of a neural network. The network is trained to minimize a composite loss that includes not only a data-fitting term but also a term for the PDE residual—the degree to which the network's output violates the governing physical laws. This is achieved by leveraging automatic differentiation to compute the necessary derivatives of the network's output with respect to its inputs. This powerful paradigm allows for solving both forward problems (predicting the solution to a PDE) and inverse problems (identifying unknown parameters in a PDE from data).[57]

3.3.2 PINNs for Navier-Stokes Equations

- **Key Papers and High-Quality Implementations**
  - **Paper:** Eivazi, H., Tahani, M., Schlatter, P., & Vinuesa, R. (2022). "Physics-informed neural networks for solving Reynolds-averaged Navier–Stokes equations". *Physics of Fluids*.
  - **High-Quality Implementation (PyTorch):** https://github.com/Shengfeng233/PINN-for-turbulence
  - **Significance:** This repository provides a dedicated PyTorch implementation for using PINNs to solve fluid dynamics problems, including the incompressible Navier-Stokes (NS) equations.[59] The project's internal document [1] provides an excellent PyTorch code snippet demonstrating the correct way to compute the 3D NS residuals using torch.autograd.grad with create_graph=True to enable higher-order derivatives. The linked repository complements this by providing a full project structure for a similar task (2D turbulence). It also contains a critical insight for training stability: for 2D incompressible flow, introducing a stream function to automatically satisfy the continuity equation greatly simplifies training.[59] While the project's problem is 3D, this highlights the importance of exploring clever mathematical reformulations to aid the optimizer.
  - **Training Stability:** The project's planning document [1] correctly identifies that training PINNs is notoriously challenging due to non-convex loss landscapes and stiff PDEs. It astutely references state-of-the-art mitigation strategies that are crucial for success. The most critical of these is the use of **double-precision (FP64) arithmetic** for all PINN-related computations. Recent work has demonstrated that many so-called "failure modes" are not inescapable local minima but are "precision-induced stalls" where second-order optimizers halt prematurely due to floating-point inaccuracies.[1]

Another vital strategy is the use of
adaptive sampling like the R3 (Retain-Resample-Release) algorithm to focus collocation points in regions of the domain with high PDE residuals, making training more efficient and effective.[1]

3.3.3 Contextual Surveys on Physics-Informed Machine Learning

- **Paper 1:** Various Authors. (2024). "Physics-Informed Neural Networks (PINNs): A Comprehensive Review of Methodologies, Applications, and Future Directions". *Applied Sciences*.
- **Paper 2:** Various Authors. (2024). "A Review of Physics-Informed Neural Networks". *ResearchGate*.
- **Significance:** These recent reviews provide a broad and up-to-date overview of the rapidly evolving PINN landscape.[64] They cover the fundamental concepts, architectural variations (e.g., domain decomposition, hybrid models), theoretical underpinnings (convergence analysis, generalization guarantees), and common challenges. A key challenge that is consistently highlighted across the literature is the difficulty of training, stemming from the highly non-convex and often competitive nature of the multi-term loss function.[66] This reinforces the importance of the advanced mitigation strategies (FP64 precision, adaptive sampling, curriculum learning) identified in the project blueprint as non-negotiable requirements for successful implementation.

3.4 Loss Functions for High-Performance Classification

The blueprint selects two specialized loss functions, Asymmetric Loss and AUC Margin Loss, to directly address the challenges of class imbalance and to optimize for the primary competition metric. This reflects a sophisticated, metric-driven approach to model training.

3.4.1 Asymmetric Loss (ASL)

- **Primary Foundational Paper(s)**
    - **Paper:** Ridnik, T., Ben-Baruch, E., Zamir, N., Noy, A., Friedman, I., Protter, M., & Zelnik-Manor, L. (2021). "Asymmetric Loss for Multi-Label Classification". In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.
    - **Significance:** This paper introduces **Asymmetric Loss (ASL)**, a novel loss function designed to address the severe positive-negative imbalance that is common in multi-label classification and other real-world datasets.[67] ASL is a modification of the popular Focal Loss but introduces two key innovations: (1) **Asymmetric Focusing**, which decouples the focusing parameters for positive ($\gamma+$) and negative ($\gamma-$) samples. This allows for different exponential decay rates, enabling the model to down-weight easy negatives ($p \rightarrow 0$) heavily (high $\gamma-$) while maintaining a strong gradient signal from the rare positive samples (low $\gamma+$). (2) **Asymmetric Probability Shifting**, which uses a probability margin to hard-threshold and completely discard the loss contribution from very easy negative samples.[67] Together, these mechanisms allow the model to focus training on hard negative examples without letting the sheer number of negatives overwhelm the learning signal from the few positive examples.[68]
- **High-Quality Implementations**
    - **Official Repository (PyTorch):** https://github.com/Alibaba-MIIL/ASL
    - **Analysis:** This is the official PyTorch implementation from the authors at Alibaba Group.[67] The repository provides a clean, well-documented, and easy-to-integrate implementation of the ASL function. It is the definitive source for this loss function. The continued relevance and evolution of this approach are demonstrated by follow-up work such as "Robust Asymmetric Loss" [69], which builds upon ASL for long-tailed learning and also provides a public GitHub repository.

3.4.2 AUC Margin Loss

- **Primary Foundational Paper(s)**
    - **Paper:** Yuan, Z., et al. (2021). "Large-Scale Robust Deep AUC Maximization: A New Surrogate Loss and Empirical Studies on Medical Image Classification". In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.
    - **Significance:** This paper tackles the significant challenge of directly optimizing the Area Under the ROC Curve (AUC), which is a standard

  evaluation metric for binary classification, especially on imbalanced datasets.[70] Direct optimization is difficult because AUC is a non-decomposable, pairwise metric that depends on the entire dataset. The authors propose a new
  AUC Margin Loss, a min-max surrogate loss function designed to be more robust to noisy data and less sensitive to easy negative examples compared to the previous state-of-the-art AUC Square Loss.[71] The formulation as a min-max objective maintains scalability for training deep neural networks on large datasets, a problem that plagued earlier pairwise approaches.[73]

- **High-Quality Implementations**
  - **Official Library (PyTorch):** https://github.com/Optimization-AI/LibAUC
  - **Analysis:** The authors have released their work as a dedicated, open-source PyTorch library named LibAUC.[72] This library provides the necessary components for this technique, including the
    AUCM_Loss class and a corresponding custom optimizer (PESG or Adam) that is required to handle the min-max optimization procedure. The library is well-documented, actively maintained, and is the definitive, high-quality implementation for directly optimizing AUC in PyTorch.[73]

3.5 Loss Balancing: Uncertainty Weighting

- **Primary Foundational Paper(s)**
  - **Paper:** Kendall, A., Gal, Y., & Cipolla, R. (2018). "Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics". In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
  - **Significance:** This paper proposes a principled and elegant probabilistic approach to balancing multiple loss terms in a multi-task learning (MTL) setting.[75] Instead of relying on a tedious and often suboptimal manual grid search for static loss weights, the method proposes to weigh each task's loss based on its
    homoscedastic uncertainty—a measure of the task's inherent, observation-independent noise. In practice, this is implemented by having the model predict a learnable log-variance term, log(σi2), for each of the i tasks. The total loss is then formulated as a sum over the tasks, with each task's loss Li being down-weighted by its variance and regularized by the log-variance term: Ltotal=Σi2σi21Li+21log(σi2). The model learns to balance the tasks by

adjusting the σi parameters automatically during training.[1]

- **Strategic Rationale:** The blueprint's Expert Classifier is trained with a composite loss function containing at least four distinct terms (Lclass, Lseg, Lloc, Lphysics). Manually tuning the weights for these terms would be intractable and likely lead to a suboptimal balance. The internal project document correctly identifies Uncertainty Weighting (UW) as the ideal solution, particularly for a computationally constrained environment like a Kaggle competition. Its key advantage is its negligible computational overhead, as it only adds one learnable scalar parameter per task and requires no complex gradient manipulations or multiple backward passes, unlike more expensive gradient-based methods like MGDA.[1] This is a critical choice for ensuring the feasibility and performance of the multi-task training strategy within the competition's runtime limits.

- **High-Quality Implementations**
  - **PyTorch Repository:** https://github.com/Mikoto10032/AutomaticWeightedLoss
  - **Analysis:** This repository provides a clean, self-contained, and easy-to-use PyTorch implementation of the uncertainty weighting scheme.[77] It wraps the entire logic in a simple AutomaticWeightedLoss module that takes multiple loss tensors as input and returns the final, weighted scalar sum. The repository's documentation also correctly notes that its implementation is based on a follow-up paper that improves upon the original formulation to prevent the total loss from becoming negative during training, making this a robust and modern implementation of the concept.[77]

Part IV: Final Ensemble & Submission

4.1 Ensemble Method: Stacking

- **Primary Foundational Paper(s)**
  - **Paper:** Wolpert, D. H. (1992). "Stacked Generalization". *Neural Networks*.
  - **Significance:** This is the seminal paper that introduced the concept of

Stacked Generalization, now commonly known as "stacking".[78] The core idea is to move beyond simple ensemble methods like averaging or voting. Stacking uses the predictions of multiple diverse base models (referred to as level-0 generalizers) as input features for a second-level "meta-model" (the level-1 generalizer). This meta-model then learns how to optimally combine the base models' predictions to produce a final, often more accurate, output. Wolpert framed this as a means of "non-linearly combining generalizers to make a new generalizer," which can capture more complex relationships between the base model outputs than a simple linear combination.[78]

  - **Strategic Rationale:** The project blueprint proposes a highly novel and sophisticated application of the stacking principle. Instead of using only the standard prediction probabilities from the base models as meta-features, the plan is to augment this feature set with unique, interpretable metrics derived from the custom architecture of the Expert Classifier. These include the **AKOrN order parameter** (a measure of the internal "consensus" or synchronization of the oscillatory neurons) and the **PINN physics loss** (a measure of the physical plausibility of the model's learned representation of blood flow for a given candidate).[1] This approach transforms the meta-learner's task. It is no longer just learning to weigh predictions; it is learning to weigh predictions based on the base models' internal state and reliability. This gives the level-2 XGBoost model much richer information, allowing it to learn complex rules such as "distrust predictions from model 3 if its PINN loss is high, even if its predicted probability is also high".[1] This is a key innovation that leverages the unique interpretability of the proposed architecture to create a more intelligent and robust final ensemble.

- **High-Quality Implementations**
  - **Scikit-learn + XGBoost Tutorials:**
    - https://machinelearningmastery.com/stacking-ensemble-machine-learning-with-python/
    - https://xgboosting.com/stacking-ensemble-with-xgboost-meta-model-final-model/
  - **Analysis:** Stacking is a high-level algorithmic concept, and its implementation is made straightforward by modern machine learning libraries. The scikit-learn library provides dedicated StackingClassifier and StackingRegressor classes that formalize the entire process, including the cross-validation-based generation of the level-1 training data.[80] The linked tutorials provide clear, idiomatic Python code demonstrating exactly how to define a list of base estimators (e.g.,
RandomForestClassifier, SVC) and specify a final_estimator, which in the

project's case would be an XGBClassifier.[80] These guides provide the exact code template needed to implement the level-2 meta-learner.

- **Recent Survey/Review Papers**
  - **Paper 1:** Dietterich, T. G. (2000). "Ensemble Methods in Machine Learning". In *Multiple Classifier Systems*.
  - **Paper 2:** Sagi, O., & Rokach, L. (2018). "Ensemble learning: A survey". *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*.
  - **Significance:** While older, Dietterich's paper [84] is a foundational review that provides a clear theoretical explanation for *why* ensemble methods work. It attributes their success to a combination of statistical (reducing variance by averaging over multiple hypotheses), computational (escaping local minima), and representational (expanding the space of representable functions) advantages. More recent surveys [85] categorize the main families of ensemble methods (bagging, boosting, and stacking) and discuss their properties. They consistently reinforce the key principle that ensembles work best when the base classifiers are both accurate and diverse—that is, they tend to make different errors. The blueprint's novel use of interpretability metrics (order parameter, physics loss) as meta-features is a sophisticated and direct way to explicitly model and leverage this diversity and reliability in the meta-learning stage.

Part V: Data Harmonization & Preprocessing

5.1 Intensity Normalization: Nyúl Histogram Matching

- **Primary Foundational Paper(s)**
  - **Paper:** Nyúl, L. G., Udupa, J. K., & Zhang, X. (2000). "New variants of a method of MRI scale standardization". *IEEE Transactions on Medical Imaging*.
  - **Significance:** This paper addresses a fundamental and persistent problem in magnetic resonance imaging: the lack of a standard, quantifiable intensity scale. Unlike in CT, where Hounsfield units have a direct physical meaning, MR image intensities can vary significantly across patients, scanners, and even different scan sessions for the same patient due to scanner-dependent

variations.[87] The paper proposes a robust two-step method to standardize these intensities. The method is based on piecewise linear histogram matching, where it first learns a "standard" histogram from a training set by mapping specific intensity percentile landmarks (e.g., deciles) to a predefined standard scale. In the second step, new images are normalized by mapping their corresponding intensity percentiles to this learned standard histogram.[61]

- **High-Quality Implementations**
  - **Python Library:** intensity-normalization
  - **Analysis:** The intensity-normalization Python package provides a direct, robust, and easy-to-use implementation of the Nyúl and Udupa method. It is available as the NyulNormalize class.[61] The library's documentation is excellent, providing clear instructions for both a Python API and a command-line interface. Crucially, it includes methods to fit a standard histogram from a set of training images, normalize new images against that standard, and save and load the learned histogram to a file. This last feature is essential for ensuring reproducible processing across all stages of the project (training, validation, and inference).[61]
- **Recent Survey/Review Papers**
  - **Paper 1:** Jansen, R. (2019). "Evaluation of MRI intensity normalization methods on images with pathology". *Master's Thesis, University of Twente*.
  - **Paper 2:** Shah, V., et al. (2025). "Systematic comparison of common intensity normalization methods in multi-parametric magnetic resonance imaging of cancer". *Frontiers in Oncology*.
  - **Significance:** These reviews provide valuable comparative analyses of various MRI intensity normalization techniques, including histogram-based methods like Nyúl's, simpler statistical methods like Z-score normalization, and tissue-based methods.[88] A critical insight that emerges from these studies, particularly Jansen (2019), is that the performance of many normalization methods can be significantly affected by the presence of pathology (e.g., white matter lesions), which can distort the image histogram and bias the landmark estimation.[88] This is a critical consideration for the aneurysm detection project, as the target pathology and potential co-morbidities will be present in the data. While some studies find that Z-score normalization performed within a specific organ mask can be highly robust [89], Nyúl's method remains a standard and powerful technique for harmonizing intensities across a large and diverse cohort.

5.2 Brain Extraction: SynthStrip

- **Primary Foundational Paper(s)**
  - **Paper:** Hoopes, A., Mora, J. S., Dalca, A. V., Fischl, B., & Hoffmann, M. (2022). "SynthStrip: Skull-Stripping for Any Brain Image". *NeuroImage*.
  - **Significance:** This paper introduces **SynthStrip**, a learning-based skull-stripping (brain extraction) tool designed for exceptional robustness and generalization across a wide variety of imaging data.[91] Its core innovation lies not in the network architecture itself, but in its unique training strategy. Instead of being trained on a curated dataset of real images, SynthStrip is trained entirely on a massive and diverse dataset of *synthetic* images. These images are generated by taking anatomical label maps and assigning random intensities, deformations, and artifacts. By generating a training set with a range of appearances that far exceeds what is seen in real-world data, the resulting model learns to be agnostic to the specific acquisition parameters (e.g., modality, contrast, resolution) of any given input image.[91]
- **Official Tool and Implementation**
  - **FreeSurfer Integration & Container:** https://surfer.nmr.mgh.harvard.edu/docs/synthstrip/
  - **Analysis:** SynthStrip is distributed as a robust, pre-trained, and easy-to-use command-line tool, making it ideal for integration into a data processing pipeline.[94] It is officially integrated into the widely-used FreeSurfer neuroimaging suite (as the mri_synthstrip command) and is also available as a standalone container (e.g., Docker, Singularity) for users who do not wish to install the full FreeSurfer package. The official website provides clear instructions for usage, the pre-trained model weights, and even the full synthetic dataset that was used for training, ensuring complete transparency and utility for the project.[93]
- **Recent Survey/Review Papers**
  - **Context:** The foundational SynthStrip paper itself contains an excellent and comprehensive review of prior skull-stripping methods.[91] It categorizes these methods into classical approaches (e.g., watershed algorithms, deformable surfaces like those used in the popular BET tool) and more recent learning-based approaches. The paper clearly articulates the primary limitation of nearly all previous tools: they are often tailored to specific image types (e.g., high-resolution, T1-weighted research scans) and tend to perform poorly or fail catastrophically on other contrasts or on clinical-quality

acquisitions (e.g., thick-slice scans).[91] SynthStrip was designed specifically to overcome this critical limitation, making it the ideal choice for a project that will likely encounter data from diverse clinical sources with varying acquisition protocols.

Table 1: Comparison of Self-Supervised Pre-training Backbones

Backbone	Core Principle	Key Innovation	Strategic Advantage for This Project
SparK	Generative Masked Image Modeling (MIM)	Adapts MIM to standard ConvNets using sparse convolutions to efficiently process only unmasked regions.	**Robustness:** Proven to maintain high performance even when fine-tuning on scarce data, de-risking the model against rare aneurysm subtypes in the unseen test set.
WaveFormer	Hierarchical Transformer with Wavelet Transform	Applies self-attention only to the compact, low-frequency component of feature maps decomposed by a Discrete Wavelet Transform (DWT).	**Efficiency:** Drastically reduces the computational and memory cost of processing 3D volumes while retaining the ability to model long-range dependencies, crucial for a computationally intensive pipeline.

Table 2: Overview of Loss Functions for Candidate Classification

Loss Function	Foundational	Core Problem	Key Mechanism	Strategic

	Paper	Addressed		Advantage
Asymmetric Loss (ASL)	Ridnik et al. (2021)	Severe positive-negative class imbalance.	**Asymmetric Focusing:** Decouples the focusing parameters ($\gamma+,\gamma-$) to heavily down-weight easy negatives while preserving gradients from rare positives.	**Balanced Learning:** Prevents the vast number of negative candidates from overwhelming the training signal from the few true positive aneurysm candidates.
AUC Margin Loss	Yuan et al. (2021)	Direct optimization of the AUC score, a non-decomposable metric.	**Min-Max Surrogate Loss:** Provides a robust, margin-based surrogate for the AUC score that is scalable for deep learning and less sensitive to noisy or easy examples.	**Direct Metric Optimization:** Aligns the training objective directly with the primary competition evaluation metric, which is crucial for maximizing performance on imbalanced datasets.

Table 3: Summary of Data Preprocessing and Harmonization Tools

Tool	Foundational Paper	Core Function	Key Advantage	Implementation
Nyúl Histogram Matching	Nyúl et al. (2000)	MRI Intensity Standardization	Maps image intensity percentiles to a learned	**Robust Standardization:** Ensures that similar tissue

			standard histogram, creating consistent intensity scales across a diverse cohort of scans.	types have similar intensity values, which is critical for the stability of downstream models.
SynthStrip	Hoopes et al. (2022)	Skull-Stripping (Brain Extraction)	A deep learning model trained entirely on synthetic data, enabling it to generalize across a vast range of modalities, contrasts, and resolutions.	**Modality-Agnostic:** Provides highly accurate brain masks for any input image type (MRI, CT, etc.) without needing to be retrained, ensuring robust preprocessing for diverse datasets.

Conclusion

The "Definitive Master Blueprint" outlines a remarkably sophisticated and strategically coherent plan for the RSNA 2025 Intracranial Aneurysm Detection challenge. The selection of components demonstrates a deep understanding of the problem domain and a forward-thinking approach to model design. Several key themes emerge from the analysis of its foundational resources:

1. **A Unified Philosophy of Robustness:** The blueprint is not a mere collection of state-of-the-art techniques but is built on a consistent philosophy of resilience. The choices of SparK for pre-training (robust to data scarcity), Tversky Loss for segmentation (prioritizing recall to avoid catastrophic failures), and both Asymmetric Loss and AUC Margin Loss for classification (explicitly handling class imbalance) form a defensive bulwark against the most likely failure modes in a high-stakes competition with an unknown, imbalanced test set.
2. **Physics and Dynamics as Inductive Priors:** The integration of Artificial Kuramoto Oscillatory Neurons (AKOrN) and a Physics-Informed Neural Network

(PINN) head represents a paradigm shift from pure data-driven pattern recognition. AKOrN introduces principles of dynamical systems to encourage the "binding" of related features, while the PINN regularizer enforces consistency with the physical laws of fluid dynamics. Together, these components inject powerful, domain-specific inductive biases into the model, compelling it to learn a physically plausible latent space that should lead to superior generalization.
3. **Interpretability as a Learnable Feature:** The proposed stacking ensemble is a significant innovation. By using the AKOrN order parameter and the PINN physics loss as meta-features, the ensemble is designed to learn not just *what* the base models predict, but *how reliably* they arrive at those predictions. This transforms the ensemble from a simple aggregator into an intelligent arbiter that can leverage the unique, interpretable byproducts of its custom architecture to make more robust final decisions.

Based on this comprehensive review, one primary actionable recommendation emerges:

- **Actionable Recommendation: Prioritize WaveFormer Implementation:** There is a critical implementation gap for the WaveFormer architecture. The most prominent public repositories are for an unrelated audio processing task. The project plan must therefore allocate dedicated engineering resources for an in-house implementation of the 3D medical WaveFormer, based on the detailed architectural specifications in the Perera et al. (2024) paper.

In summary, the blueprint represents a state-of-the-art, well-researched, and strategically sound approach. Its successful execution, contingent on addressing the identified implementation gap, positions the project for a highly competitive entry in the RSNA 2025 challenge.

Works cited

1. A Physics.pdf
2. SparK — MMPretrain 1.2.0 documentation - Read the Docs, accessed August 11, 2025, https://mmpretrain.readthedocs.io/en/latest/papers/spark.html
3. [Literature Review] WaveFormer: A 3D Transformer with Wavelet-Driven Feature Representation for Efficient Medical Image Segmentation - Moonlight | AI Colleague for Research Papers, accessed August 11, 2025, https://www.themoonlight.io/en/review/waveformer-a-3d-transformer-with-wavelet-driven-feature-representation-for-efficient-medical-image-segmentation
4. Revealing the Dark Secrets of Masked Image Modeling - CVF Open Access, accessed August 11, 2025, https://openaccess.thecvf.com/content/CVPR2023/papers/Xie_Revealing_the_Dar

k_Secrets_of_Masked_Image_Modeling_CVPR_2023_paper.pdf
5. [Literature Review] Masked Image Modeling: A Survey - Moonlight, accessed August 11, 2025, https://www.themoonlight.io/en/review/masked-image-modeling-a-survey
6. (PDF) Masked Image Modeling: A Survey - ResearchGate, accessed August 11, 2025, https://www.researchgate.net/publication/393673075_Masked_Image_Modeling_A_Survey
7. arxiv.org, accessed August 11, 2025, https://arxiv.org/html/2404.15580v2
8. MiM: Mask in Mask Self-Supervised Pre-Training for 3D Medical Image Analysis - arXiv, accessed August 11, 2025, https://arxiv.org/abs/2404.15580
9. HybridMIM: A Hybrid Masked Image Modeling Framework for 3D Medical Image Segmentation - GitHub, accessed August 11, 2025, https://github.com/ge-xing/HybridMIM
10. ucasligang/awesome-MIM: Reading list for research topics in Masked Image Modeling - GitHub, accessed August 11, 2025, https://github.com/ucasligang/awesome-MIM
11. ZiyuGuo99/Awesome-MIM-1: Awesome List of Masked Image Modeling (MIM) Papers for Self-supervised Visual Representation Learning - GitHub, accessed August 11, 2025, https://github.com/ZiyuGuo99/Awesome-MIM-1
12. Masked Image Modeling Advances 3D Medical Image Analysis - CVF Open Access, accessed August 11, 2025, https://openaccess.thecvf.com/content/WACV2023/papers/Chen_Masked_Image_Modeling_Advances_3D_Medical_Image_Analysis_WACV_2023_paper.pdf
13. GMIM: Self-supervised pre-training for 3D medical image segmentation with adaptive and hierarchical masked image modeling - PubMed, accessed August 11, 2025, https://pubmed.ncbi.nlm.nih.gov/38728994/
14. Masked Image Modeling Advances 3D Medical Image Analysis | Request PDF - ResearchGate, accessed August 11, 2025, https://www.researchgate.net/publication/368314169_Masked_Image_Modeling_Advances_3D_Medical_Image_Analysis
15. WaveFormer: A 3D Transformer with Wavelet-Driven Feature Representation for Efficient Medical Image Segmentation - Powerdrill AI, accessed August 11, 2025, https://powerdrill.ai/discover/summary-waveformer-a-3d-transformer-with-wavelet-driven-cm90exfjxnl5x07pnv1woq4tm
16. WaveFormer: A 3D Transformer with Wavelet-Driven Feature Representation for Efficient Medical Image Segmentation - ResearchGate, accessed August 11, 2025, https://www.researchgate.net/publication/390354197_WaveFormer_A_3D_Transformer_with_Wavelet-Driven_Feature_Representation_for_Efficient_Medical_Image_Segmentation
17. vb000/Waveformer: A deep neural network architecture for low-latency audio processing, accessed August 11, 2025, https://github.com/vb000/Waveformer
18. bowang-lab/MedSAM: Segment Anything in Medical Images - GitHub, accessed August 11, 2025, https://github.com/bowang-lab/MedSAM
19. MedLSAM: Localize and Segment Anything Model for 3D Medical Images -

GitHub, accessed August 11, 2025, https://github.com/openmedlab/MedLSAM
20. 3d-medical-imaging-segmentation · GitHub Topics, accessed August 11, 2025, https://github.com/topics/3d-medical-imaging-segmentation
21. Deep Learning on 3D Semantic Segmentation: A Detailed Review - MDPI, accessed August 11, 2025, https://www.mdpi.com/2072-4292/17/2/298
22. Medical Image Segmentation: A Comprehensive Review of Deep Learning-Based Methods, accessed August 11, 2025, https://www.mdpi.com/2379-139X/11/5/52
23. Efficient 3D Deep Learning Model for Medical Image Semantic Segmentation, accessed August 11, 2025, https://www.researchgate.net/publication/346747284_Efficient_3D_Deep_Learning_Model_for_Medical_Image_Semantic_Segmentation
24. Semantic Segmentation of 3D Medical Images Through a Kaleidoscope: Data from the Osteoarthritis Initiative - Proceedings of Machine Learning Research, accessed August 11, 2025, https://proceedings.mlr.press/v227/woo24a.html
25. Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks - ResearchGate, accessed August 11, 2025, https://www.researchgate.net/publication/319577938_Tversky_Loss_Function_for_Image_Segmentation_Using_3D_Fully_Convolutional_Deep_Networks
26. Tversky loss - Medium, accessed August 11, 2025, https://medium.com/@saba99/tversky-loss-902f5f8cc35f
27. Loss Function Library - Keras & PyTorch - Kaggle, accessed August 11, 2025, https://www.kaggle.com/code/bigironsphere/loss-function-library-keras-pytorch
28. Loss odyssey in medical image segmentation - Sci-Hub, accessed August 11, 2025, https://sci-hub.se/downloads/2021-06-21/20/ma2021.pdf
29. Enhancing Medical Image Segmentation Based on Loss Functions Integration, accessed August 11, 2025, https://journals.ekb.eg/article_419898_0.html
30. Loss Functions in Deep Learning: A Comprehensive Review - arXiv, accessed August 11, 2025, https://arxiv.org/html/2504.04242v1
31. SegFormer3D: An Efficient Transformer for 3D Medical Image Segmentation, accessed August 11, 2025, https://ohiostate.elsevierpure.com/en/publications/segformer3d-an-efficient-transformer-for-3d-medical-image-segment
32. SegFormer3D: Efficient 3D Med Image Segmentation - Emergent Mind, accessed August 11, 2025, https://www.emergentmind.com/papers/2404.10156
33. SegFormer3D: an Efficient Transformer for 3D Medical Image ..., accessed August 11, 2025, https://www.aimodels.fyi/papers/arxiv/segformer3d-efficient-transformer-3d-medical-image-segmentation
34. OSUPCVLab/SegFormer3D: Official Implementation of ... - GitHub, accessed August 11, 2025, https://github.com/OSUPCVLab/SegFormer3D
35. Deep Vessel Segmentation by Learning Graphical Connectivity - ResearchGate, accessed August 11, 2025, https://www.researchgate.net/publication/335667210_Deep_Vessel_Segmentation_by_Learning_Graphical_Connectivity
36. Retinal vessel segmentation via deep learning network and fully-connected

conditional random fields - Semantic Scholar, accessed August 11, 2025, https://www.semanticscholar.org/paper/Retinal-vessel-segmentation-via-deep-learning-and-Fu-Xu/1f8564d458f2e36f5d1ad756c4a92485e45cf7ff
37. syshin1014/VGN: Deep Vessel Segmentation by Learning ... - GitHub, accessed August 11, 2025, https://github.com/syshin1014/VGN
38. Skeletonize — skimage 0.25.2 documentation, accessed August 11, 2025, https://scikit-image.org/docs/0.25.x/auto_examples/edges/plot_skeleton.html
39. [PDF] Graph Attention Networks - Semantic Scholar, accessed August 11, 2025, https://www.semanticscholar.org/paper/Graph-Attention-Networks-Velickovic-Cucurull/33998aff64ce51df8dee45989cdca4b6b1329ec4
40. (PDF) Graph Attention Networks (2018) | Petar Veličković | 14172 Citations - SciSpace, accessed August 11, 2025, https://scispace.com/papers/graph-attention-networks-1xed6mcmzg
41. HONGAT: Graph Attention Networks in the Presence of High-Order Neighbors, accessed August 11, 2025, https://ojs.aaai.org/index.php/AAAI/article/view/29615/31042
42. Graph Attention Networks (GAT), accessed August 11, 2025, https://nn.labml.ai/graphs/gat/index.html
43. A Comparative Study of Graph Neural Networks for Shape Classification in Neuroimaging, accessed August 11, 2025, https://proceedings.mlr.press/v194/shehata22a.html
44. Graph-Based Deep Learning for Medical Diagnosis and Analysis: Past, Present and Future - PMC - PubMed Central, accessed August 11, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC8309939/
45. (PDF) A Systematic Review of Graph Neural Network in Healthcare-Based Applications: Recent Advances, Trends, and Future Directions - ResearchGate, accessed August 11, 2025, https://www.researchgate.net/publication/377454172_A_systematic_review_of_graph_neural_network_in_healthcare-based_applications_recent_advances_trends_and_future_directions
46. [Literature Review] Artificial Kuramoto Oscillatory Neurons, accessed August 11, 2025, https://www.themoonlight.io/en/review/artificial-kuramoto-oscillatory-neurons
47. Artificial Kuramoto Oscillatory Neurons (AKOrN) - ICLR 2025 - Takeru Miyato, accessed August 11, 2025, https://takerum.github.io/akorn_project_page/
48. ARTIFICIAL KURAMOTO OSCILLATORY NEURONS - Andreas Geiger, accessed August 11, 2025, https://www.cvlibs.net/publications/Miyato2025ICLR.pdf
49. autonomousvision/akorn: [ICLR'25] Artificial Kuramoto ... - GitHub, accessed August 11, 2025, https://github.com/autonomousvision/akorn
50. DataSQRL/Acorn: Flexible Agent Framework for building Chat Bots, AI agents, and LLM microservices on the JVM. - GitHub, accessed August 11, 2025, https://github.com/DataSQRL/Acorn
51. thomasleese/acorn: Powerful and elegant programming language. - GitHub, accessed August 11, 2025, https://github.com/thomasleese/acorn
52. FiLM: Visual Reasoning with a General Conditioning Layer - ResearchGate,

accessed August 11, 2025, https://www.researchgate.net/publication/320014293_FiLM_Visual_Reasoning_with_a_General_Conditioning_Layer
53. FiLM: Visual Reasoning with a General Conditioning Layer | Request PDF - ResearchGate, accessed August 11, 2025, https://www.researchgate.net/publication/386685275_FiLM_Visual_Reasoning_with_a_General_Conditioning_Layer
54. FiLM- Visual Reasoning with a General Conditioning Layer - ML Retrospectives, accessed August 11, 2025, https://ml-retrospectives.github.io/neurips2019/accepted_retrospectives/2019/film/
55. torch_geometric.nn.conv.FiLMConv — pytorch_geometric ..., accessed August 11, 2025, https://pytorch-geometric.readthedocs.io/en/latest/generated/torch_geometric.nn.conv.FiLMConv.html
56. GNN-FiLM: Graph Neural Networks with Feature-wise Linear Modulation, accessed August 11, 2025, https://proceedings.mlr.press/v119/brockschmidt20a.html
57. Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations | Request PDF - ResearchGate, accessed August 11, 2025, https://www.researchgate.net/publication/328720075_Physics-Informed_Neural_Networks_A_Deep_Learning_Framework_for_Solving_Forward_and_Inverse_Problems_Involving_Nonlinear_Partial_Differential_Equations
58. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations (Journal Article) | OSTI.GOV, accessed August 11, 2025, https://www.osti.gov/pages/biblio/1595805
59. Shengfeng233/PINN-for-turbulence: A pytorch ... - GitHub, accessed August 11, 2025, https://github.com/Shengfeng233/PINN-for-turbulence
60. Physics-informed neural networks for analyzing size effect and identifying parameters in piezoelectric semiconductor nanowires - AIP Publishing, accessed August 11, 2025, https://pubs.aip.org/aip/jap/article/137/2/024303/3330269/Physics-informed-neural-networks-for-analyzing
61. intensity-normalization Documentation - Read the Docs, accessed August 11, 2025, https://readthedocs.org/projects/intensity-normalization/downloads/pdf/latest/
62. HySparK: Hybrid Sparse Masking for Large Scale Medical Image Pre-Training - MICCAI, accessed August 11, 2025, https://papers.miccai.org/miccai-2024/395-Paper0897.html
63. MiM: Mask in Mask Self-Supervised Pre-Training for 3D Medical Image Analysis | Request PDF - ResearchGate, accessed August 11, 2025, https://www.researchgate.net/publication/391161659_MiM_Mask_in_Mask_Self-Supervised_Pre-Training_for_3D_Medical_Image_Analysis

64. Physics-Informed Neural Networks: A Review of Methodological Evolution, Theoretical Foundations, and Interdisciplinary Frontiers Toward Next-Generation Scientific Computing - MDPI, accessed August 11, 2025, https://www.mdpi.com/2076-3417/15/14/8092
65. Understanding Physics-Informed Neural Networks: Techniques, Applications, Trends, and Challenges - MDPI, accessed August 11, 2025, https://www.mdpi.com/2673-2688/5/3/74
66. (PDF) A Review of Physics-Informed Neural Networks - ResearchGate, accessed August 11, 2025, https://www.researchgate.net/publication/388357372_A_Review_of_Physics-Informed_Neural_Networks
67. Asymmetric Loss for Multi-Label Classification - CVF Open Access, accessed August 11, 2025, https://openaccess.thecvf.com/content/ICCV2021/papers/Ridnik_Asymmetric_Loss_for_Multi-Label_Classification_ICCV_2021_paper.pdf
68. Asymmetric Loss For Multi-Label Classification | Request PDF - ResearchGate, accessed August 11, 2025, https://www.researchgate.net/publication/344422815_Asymmetric_Loss_For_Multi-Label_Classification
69. Robust Asymmetric Loss for Multi-Label Long-Tailed Learning - CVF Open Access, accessed August 11, 2025, https://openaccess.thecvf.com/content/ICCV2023W/CVAMD/papers/Park_Robust_Asymmetric_Loss_for_Multi-Label_Long-Tailed_Learning_ICCVW_2023_paper.pdf
70. Deep AUC maximization - Iowa Research Online, accessed August 11, 2025, https://iro.uiowa.edu/view/pdfCoverPage?instCode=01IOWA_INST&filePid=13889111870002771&download=true
71. Large-Scale Robust Deep AUC Maximization: A New Surrogate Loss and Empirical Studies on Medical Image Classification - CVF Open Access, accessed August 11, 2025, https://openaccess.thecvf.com/content/ICCV2021/papers/Yuan_Large-Scale_Robust_Deep_AUC_Maximization_A_New_Surrogate_Loss_and_ICCV_2021_paper.pdf
72. (PDF) Robust Deep AUC Maximization: A New Surrogate Loss and Empirical Studies on Medical Image Classification - ResearchGate, accessed August 11, 2025, https://www.researchgate.net/publication/346701856_Robust_Deep_AUC_Maximization_A_New_Surrogate_Loss_and_Empirical_Studies_on_Medical_Image_Classification
73. (PDF) AUC margin loss for limited, imbalanced and noisy medical ..., accessed August 11, 2025, https://www.researchgate.net/publication/374081308_AUC_margin_loss_for_limited_imbalanced_and_noisy_medical_image_diagnosis_-_a_case_study_on_CheXpert5000
74. Optimization-AI/LibAUC: LibAUC: A Deep Learning Library ... - GitHub, accessed

August 11, 2025, https://github.com/Optimization-AI/LibAUC
75. Multi-task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics, accessed August 11, 2025, https://www.repository.cam.ac.uk/items/372e93bb-ef37-4733-b8e7-5a26080f1fb4
76. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics, accessed August 11, 2025, https://core.ac.uk/outputs/346585668/?source=1&algorithmId=15&similarToDoc=&similarToDocKey=URL&recSetID=3490bc08-d2bd-463e-beb9-bbe4fb933591&position=1&recommendation_type=same_repo&otherRecs=346585668%2C482248167%2C211994099%2C287545265%2C482251373
77. Mikoto10032/AutomaticWeightedLoss: Multi-task learning ... - GitHub, accessed August 11, 2025, https://github.com/Mikoto10032/AutomaticWeightedLoss
78. Stacked Generalization - Data Science For The Greater Good, accessed August 11, 2025, https://kpokrass.github.io/stacked_generalization
79. Stacked Generalization: when does it work? - IJCAI, accessed August 11, 2025, https://www.ijcai.org/Proceedings/97-2/Papers/011.pdf
80. Stacking Ensemble Machine Learning With Python - MachineLearningMastery.com, accessed August 11, 2025, https://machinelearningmastery.com/stacking-ensemble-machine-learning-with-python/
81. 1.11. Ensembles: Gradient boosting, random forests, bagging, voting, stacking - Scikit-learn, accessed August 11, 2025, https://scikit-learn.org/stable/modules/ensemble.html
82. Stacking Ensemble With XGBoost Meta Model (Final Model), accessed August 11, 2025, https://xgboosting.com/stacking-ensemble-with-xgboost-meta-model-final-model/
83. Understanding Stacking Ensemble Method: A Comprehensive Guide -PART IV, accessed August 11, 2025, https://dev.to/sreeni5018/understanding-stacking-ensemble-method-a-comprehensive-guide-part-iv-1c29
84. Ensemble Methods in Machine Learning - College of Engineering | Oregon State University, accessed August 11, 2025, https://web.engr.oregonstate.edu/~tgd/publications/mcs-ensembles.pdf
85. A Review on Ensemble Learning Methods: Machine Learning Approach - ijrpr, accessed August 11, 2025, https://ijrpr.com/uploads/V6ISSUE2/IJRPR38994.pdf
86. (PDF) A Review on Ensemble Learning Methods: Machine Learning Approach, accessed August 11, 2025, https://www.researchgate.net/publication/389529749_A_Review_on_Ensemble_Learning_Methods_Machine_Learning_Approach
87. New variants of a method of MRI scale standardization - PubMed, accessed August 11, 2025, https://pubmed.ncbi.nlm.nih.gov/10784285/
88. Evaluation of intensity normalization methods for MR images, accessed August 11, 2025, http://essay.utwente.nl/93885/1/Jansen_MA_ET.pdf

89. Comparison of intensity normalization methods in prostate, brain, and breast cancer multi-parametric magnetic resonance imaging - Frontiers, accessed August 11, 2025, https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2025.1433444/full
90. Brain MR image intensity normalization in the presence of pathology - ISMRM, accessed August 11, 2025, https://cds.ismrm.org/protected/17MProceedings/PDFfiles/0183.html
91. SynthStrip: Skull-Stripping for Any Brain Image | Request PDF - ResearchGate, accessed August 11, 2025, https://www.researchgate.net/publication/361978194_SynthStrip_Skull-Stripping_for_Any_Brain_Image
92. [2203.09974] SynthStrip: Skull-Stripping for Any Brain Image - arXiv, accessed August 11, 2025, https://arxiv.org/abs/2203.09974
93. (ISMRM 2022) SynthStrip: skull stripping for any brain image, accessed August 11, 2025, https://archive.ismrm.org/2022/2686.html
94. SynthStrip: Skull-Stripping for Any Brain Image - FreeSurfer, accessed August 11, 2025, https://surfer.nmr.mgh.harvard.edu/docs/synthstrip/
95. SynthStrip Skull Stripping For Any Brain Image - 2022 - NeuroImage | PDF - Scribd, accessed August 11, 2025, https://www.scribd.com/document/763233336/SynthStrip-skull-stripping-for-any-brain-image-2022-NeuroImage