

Deep Learning-Based Prediction of Asphalt Mixture Mechanical Properties from Aggregate Morphology and Mesostructural Features

Fitan Deshpande

Department of Electrical Engineering and Computer Science, University of Missouri,
Columbia, MO, USA.
fitand@missouri.edu

Kirk Hays

Department of Computer Science, University of North Texas, Denton, TX, USA.
kirk.work@unt.edu

Abstract

Asphalt mixtures are complex particulate composites whose mechanical behavior emerges from interactions among aggregate shape, gradation, binder properties, air-void structure, and compaction history. Conventional mixture design and performance prediction often rely on empirical correlations and laboratory testing, which can be costly, time-consuming, and insufficiently sensitive to mesostructural variability. This paper examines the system-level potential of deep learning to predict asphalt mixture mechanical properties from aggregate morphology and mesostructural features. It argues that accurate prediction is not merely a model-fitting problem but a socio-technical infrastructure problem involving imaging pipelines, dataset governance, uncertainty quantification, deployment workflows, and sustainability-oriented decision making. The discussion synthesizes advances in aggregate shape characterization, image-based mesostructure analysis, convolutional and attention-based neural architectures, explainability methods, and reliability engineering. Particular attention is given to structural trade-offs among accuracy, interpretability, computational cost, and field deployability. The paper also considers fairness and robustness when models are transferred across quarries, climates, binder sources, and mixture designs. It concludes that deep learning can augment asphalt mixture design if integrated with material standards, model documentation, calibrated uncertainty, and continuous monitoring. Without such governance, predictive systems risk becoming brittle black boxes that obscure rather than illuminate the physical mechanisms governing pavement performance. The proposed perspective positions deep learning as one component within a broader materials informatics ecosystem that connects aggregate imaging, laboratory validation, construction quality control, and long-term pavement management.

Keywords

asphalt mixture, aggregate morphology, mesostructure, deep learning, pavement engineering, uncertainty quantification, infrastructure governance, sustainability.

1. Introduction

Asphalt mixtures remain among the most widely used materials in road infrastructure, yet their mechanical performance is difficult to predict because it depends on coupled physical processes across multiple scales. At the aggregate scale, particle shape, angularity, surface

texture, and size distribution influence packing, interlock, and load transfer. At the binder scale, viscoelastic and adhesive properties govern stiffness, fatigue resistance, and moisture susceptibility. At the mixture scale, air-void distribution, aggregate orientation, and compaction energy determine the effective response under traffic and environmental loading. Traditional mixture design has relied on volumetric proportions and empirical performance tests, as documented in foundational asphalt technology literature [1]. Although these approaches are practical, they often compress rich mesostructural information into a small number of index parameters. This compression can hide variability that later manifests as rutting, cracking, or moisture damage.

Recent progress in imaging, computing, and machine learning has created new possibilities for extracting and using mesostructural information. Image analysis of asphalt concrete has long demonstrated that internal structure can be quantified rather than merely described [2]. Advances in aggregate imaging have improved measurement of angularity and texture [3], while geometric analysis has provided more precise descriptors of particle form [4]. Three-dimensional scanning and laser-based methods have further expanded the ability to capture aggregate shape characteristics [5]. These developments matter because deep learning models require informative representations, and aggregate morphology is one of the most physically meaningful inputs available before a mixture is produced.

Deep learning offers representational power that classical regression may lack. Neural networks can learn hierarchical features from images, point clouds, and tabular descriptors, and they have transformed fields such as computer vision and natural language processing [6]. Convolutional networks demonstrated the value of learned spatial filters for image classification [7], and subsequent work has expanded architectural diversity. In pavement engineering, deep learning has been applied to distress detection, material characterization, and performance prediction. However, applying these methods to asphalt mixture mechanical properties introduces distinctive challenges. The dataset is often small relative to model capacity, the physical mechanisms are multiscale, and field conditions differ from laboratory conditions. A recent study using image recognition to relate coarse aggregate morphology to asphalt mixture mesostructure illustrates the growing interest in this direction [8]. The central question is not whether deep learning can achieve low error on a held-out laboratory dataset, but whether it can support reliable decisions in design, construction, and asset management.

This paper develops a system-level analysis of deep learning-based prediction of asphalt mixture mechanical properties from aggregate morphology and mesostructural features. It emphasizes architecture, data infrastructure, uncertainty, fairness, robustness, sustainability, and governance. The argument is that predictive performance must be embedded within institutional workflows and validated against physical understanding. Without such integration, models may produce accurate-looking outputs that are not actionable or trustworthy.

2. Aggregate Morphology and Mesostructural Representation

Aggregate morphology is a broad construct encompassing form, angularity, and surface texture. Form describes the overall proportions of a particle, such as whether it is spherical, elongated, or flat. Angularity describes the sharpness of edges and corners, which affects interlocking and resistance to permanent deformation. Surface texture describes finer-scale roughness that influences friction and binder adhesion. These characteristics are not independent; a particle can be angular but rounded at a finer scale, or smooth but highly irregular in overall shape. Imaging methods have sought to quantify these dimensions using

two-dimensional projections, three-dimensional surface reconstructions, and laser scanning [3], [4], [5]. For deep learning, the challenge is to convert these physical properties into representations that retain predictive information without overwhelming the model with noise.

Mesostructural features extend beyond individual particles to the arrangement of aggregates, binder, and voids. X-ray computed tomography and related techniques allow non-destructive observation of internal structure, including air-void distribution and aggregate orientation [20]. These features mediate mechanical performance. For example, a mixture with well-graded, angular aggregates may develop strong interlock, but poor compaction can create connected voids that reduce durability. Conversely, a mixture with rounded aggregates may compact easily but exhibit lower rutting resistance. A model that predicts stiffness or fatigue life from aggregate morphology alone may therefore miss critical interactions with compaction and binder content. Deep learning can address this by learning joint representations from images of particles and images of compacted specimens, but only if the training data capture the relevant variation.

The representational choice is a structural trade-off. Hand-crafted descriptors such as angularity indices are interpretable and computationally efficient, but they may discard information relevant to complex failure modes. Learned embeddings from convolutional networks can capture subtle texture and shape patterns, but they are less transparent and may overfit to scanner-specific artifacts. Hybrid approaches that combine physically meaningful descriptors with learned features often provide a pragmatic balance. For instance, a model might use particle size distribution, angularity, and texture as explicit inputs while using a convolutional branch to extract features from aggregate images. The explicit inputs anchor the model in established material science, while the learned branch captures residual variability. This hybrid design also supports explainability because stakeholders can inspect whether predictions align with known physical trends.

Data infrastructure is as important as model architecture. Aggregate morphology datasets must document quarry source, mineralogy, crushing method, sampling protocol, and imaging conditions. Without such metadata, models cannot distinguish between true material differences and procedural artifacts. Datasheets for datasets provide a useful governance template by requiring documentation of motivation, composition, collection process, preprocessing, and intended uses [16]. For asphalt mixtures, this documentation should also include compaction method, binder grade, aging condition, and test temperature. The absence of standardized metadata is a major barrier to model transfer and reproducibility.

3. Deep Learning Architectures and Feature Integration

Deep learning architectures for asphalt mixture prediction can be organized into several families. Convolutional neural networks are well suited to image inputs such as aggregate particles, cut sections, or computed tomography slices. They learn local patterns and compose them into higher-level representations, an approach that became prominent with ImageNet classification [7]. For mesostructural images, convolutional networks can detect void shapes, aggregate contacts, and binder films. Recurrent and sequence models are less natural for spatial images but can model compaction curves or time-dependent damage evolution. Attention-based architectures, including transformers, allow long-range dependencies and flexible weighting of features [10]. They may be useful when the mechanical response depends on distant regions of a specimen, such as crack propagation pathways that cross multiple aggregate contacts.

A practical system may use multiple branches. One branch processes aggregate images to extract shape and texture embeddings. Another processes mesostructural images to extract void and contact features. A third processes tabular variables such as binder content, air-void ratio, temperature, and loading frequency. These branches are then fused through concatenation, attention, or gated integration. The fusion strategy is a key trade-off. Early fusion may allow richer interactions but can be sensitive to missing modalities. Late fusion is more robust to missing data but may miss cross-modal interactions. Intermediate fusion can balance these concerns, but it increases architectural complexity and tuning burden.

Training such models requires careful attention to data scarcity and distribution shift. Asphalt mixture datasets are often small because specimen preparation and mechanical testing are expensive. Transfer learning from large image datasets or from related pavement inspection tasks can improve sample efficiency. Data augmentation, such as rotation, scaling, and texture perturbation, can increase effective diversity, but it must respect physical constraints. Rotating an aggregate image may be acceptable, but altering gradation would change the material system. Synthetic data generation using physics-based simulation or generative models can supplement real data, but synthetic samples must be validated against experimental distributions. Otherwise, the model may learn simulator artifacts rather than material behavior.

Interpretability is not optional in infrastructure applications. A prediction of fatigue life or rutting resistance may inform expensive construction decisions or safety-critical maintenance. Explainability methods such as feature attribution can reveal which aggregate or mesostructural characteristics drive predictions [11]. However, explanations must be evaluated for stability and physical plausibility. If a model reports that surface texture is highly important but the attribution changes dramatically under small input perturbations, the explanation is unreliable. Calibration is equally important. Modern neural networks can be overconfident, and calibrated probabilities are necessary for risk-based decision making [13]. Uncertainty quantification through Bayesian approximations or ensembles can distinguish aleatoric variability from model uncertainty [12]. This distinction is essential when deciding whether to accept a prediction, request more testing, or flag a mixture for review.

4. Prediction Targets and Physical Constraints

The mechanical properties of interest include stiffness modulus, rutting resistance, fatigue life, moisture damage resistance, and fracture energy. Each target has different sensitivity to morphology and mesostructure. Stiffness is strongly influenced by binder grade, air voids, and aggregate packing. Rutting depends on aggregate interlock, angularity, and binder stiffness at high temperature. Fatigue cracking depends on binder aging, void structure, and the distribution of stress concentrations. Moisture damage depends on adhesion, cohesive strength, and the connectivity of voids. A single deep learning model could attempt multi-task prediction, but this introduces the risk of negative transfer if tasks conflict. Multi-task learning can improve generalization when tasks share representations, but it requires careful weighting and validation.

Physical constraints should be embedded in model design and evaluation. Predictions should respect monotonic trends where they are well established. For example, increasing air-void content beyond an optimal range generally reduces durability. Increasing aggregate angularity often improves rutting resistance up to a point, after which workability and compaction problems may dominate. A purely data-driven model may learn non-physical relationships if the dataset contains confounding variables. For instance, if all highly angular aggregates come from one quarry with a particular mineralogy, the model may attribute performance to

angularity when mineralogy is partly responsible. Experimental design and causal inference methods can help, but they require deliberate data collection. In practice, hybrid models that combine mechanistic simulations with neural networks may provide better extrapolation. Discrete element simulations can generate mesostructural response data, and laboratory tests can calibrate the simulator [18]. The neural network can then serve as a fast surrogate, but its training domain must be bounded by simulation validity.

Validation must go beyond random splits. Random splitting can overestimate performance because specimens from the same mixture design may appear in both training and test sets. Grouped splitting by mixture, quarry, binder source, or project is more realistic. Temporal splitting is also valuable when models are intended for deployment over time. If a model is trained on mixtures from one construction season and tested on the next, performance degradation reveals sensitivity to aging, material sourcing, and environmental variation. Reporting should include not only average error but also worst-case error, calibration, and uncertainty intervals. A model that is accurate on average but occasionally severely underestimates rutting risk is unsuitable for asset management.

5. Deployment, Governance, and Infrastructure

Deploying deep learning in asphalt mixture prediction is a socio-technical undertaking. The model must connect to imaging hardware, laboratory information management systems, mixture design software, and construction quality control platforms. This integration creates technical debt if data pipelines are fragile, undocumented, or dependent on individual researchers [14]. Technical debt in machine learning systems includes hidden feedback loops, undeclared consumers, and unstable data dependencies. In pavement engineering, a model that relies on a particular scanner calibration may fail when a different laboratory uses a different device. Governance must therefore address versioning of models, datasets, and preprocessing code.

Model documentation is a practical governance tool. Model cards can report intended use, performance across subgroups, limitations, and ethical considerations [15]. For asphalt prediction, subgroups might include aggregate types, climatic regions, traffic levels, and recycled material contents. A model card should state whether the model was trained on laboratory-compacted specimens or field cores, whether it predicts at a specific temperature, and whether it can handle recycled asphalt pavement. Datasheets for datasets complement model cards by documenting data provenance and collection conditions [16]. Together, these artifacts support accountability and informed use.

Infrastructure for deployment should include continuous monitoring. Models drift as quarries change, binder formulations evolve, and construction practices adapt. A monitoring system can track prediction residuals against periodic laboratory tests. If residuals exceed thresholds, the model can be retrained or restricted to a narrower domain. Human-in-the-loop workflows are also important. Engineers should be able to override predictions based on local knowledge. The goal is not to replace professional judgment but to augment it with probabilistic evidence. This requires interfaces that communicate uncertainty and explain key factors rather than presenting a single number.

Policy implications follow from these design choices. Public agencies may require model documentation before accepting predictive tools for mixture approval or pavement management. Standards bodies could define minimum metadata for aggregate imaging and mesostructural characterization. Procurement rules could incentivize open data sharing while

protecting proprietary quarry information. These policies must balance innovation with safety and fairness. If models are trained only on data from well-resourced laboratories, they may perform poorly in regions with different materials or testing practices. Fairness in this context means equitable model performance across geographic and economic contexts, not merely equal treatment of identical inputs.

6. Robustness, Fairness, and Sustainability

Robustness is the ability of a model to maintain performance under distribution shift, noise, and adversarial conditions. Aggregate images can be affected by lighting, dust, shadow, and segmentation errors. Mesostructural images can suffer from artifacts, resolution limits, and reconstruction errors. A robust system should include quality control for inputs and should abstain from prediction when inputs are out of distribution. Ensemble methods, data augmentation, and domain adaptation can improve robustness, but they cannot eliminate the need for validation. In safety-relevant applications, conservative uncertainty estimates are preferable to optimistic point predictions.

Fairness considerations arise when models influence resource allocation. If a pavement management system uses deep learning to prioritize maintenance, errors can disproportionately affect certain communities. A model that underestimates deterioration in low-traffic rural roads may delay repairs and increase long-term costs. A model that overestimates risk in urban neighborhoods may divert funds unnecessarily. These outcomes depend on data coverage and label quality. Agencies should audit model performance across road classes, traffic levels, and socioeconomic regions. Fairness is not a purely technical property; it is an institutional commitment to monitor and correct disparate impacts.

Sustainability adds another dimension. Asphalt production and placement consume energy and raw materials, and pavement performance affects vehicle fuel consumption and life-cycle emissions. Deep learning can support sustainability by enabling more precise mixture design, reducing overdesign, and extending pavement life. It can also help evaluate the use of recycled materials, warm-mix technologies, and alternative binders. However, machine learning itself has computational costs and environmental impacts, as highlighted in climate-oriented machine learning research [17]. The net sustainability benefit depends on whether prediction accuracy reduces material use and maintenance frequency more than it increases energy consumption. Life-cycle assessment should accompany model development. A model that improves rutting prediction by a small margin but requires enormous training compute may not be environmentally justified.

Cross-domain comparisons can clarify these trade-offs. In structural health monitoring, deep learning has been used for crack detection and damage classification [18], [19]. These applications share challenges of data scarcity, class imbalance, and domain shift. Pavement engineering can learn from their practices in sensor calibration, annotation protocols, and deployment under uncontrolled conditions. However, asphalt mixtures are distinct because the material itself is heterogeneous and the relevant features span particle to specimen scales. Predictive models must therefore integrate imaging, mechanics, and construction process data. Sustainability frameworks from pavement life-cycle assessment provide a broader evaluation lens [22]. Moisture damage testing, such as the Hamburg wheel-tracking procedure, offers an example of performance-based validation that can anchor model outputs [21].

7. Limitations and Research Agenda

Several limitations constrain current deep learning approaches. First, datasets are often proprietary and fragmented, limiting reproducibility and external validity. Second, aggregate morphology and mesostructural features are measured with different devices and protocols, creating harmonization problems. Third, mechanical tests are destructive and time-consuming, so labels are scarce. Fourth, model interpretability remains challenging, especially for multimodal architectures. Fifth, deployment requires integration with existing standards and workflows that may not be designed for probabilistic predictions. Sixth, long-term performance depends on environmental aging and traffic loading that are difficult to replicate in laboratory datasets.

A research agenda should address these limitations through shared benchmark datasets, standardized metadata, and open evaluation protocols. Federated learning could allow agencies and companies to train models without sharing proprietary data, but it raises governance questions about model ownership and auditability. Physics-informed machine learning could improve extrapolation by embedding conservation principles and material constraints. Active learning could prioritize the most informative experiments, reducing testing costs. Digital twins of pavements could provide continuous data streams for model updating, but they require sensor infrastructure and data governance. Finally, human factors research is needed to understand how engineers interpret uncertainty and explanations in practice.

8. Conclusion

Deep learning offers a powerful approach for predicting asphalt mixture mechanical properties from aggregate morphology and mesostructural features. It can learn complex relationships that traditional empirical models may miss and can integrate imaging data with laboratory and field measurements. However, the value of such systems depends on more than predictive accuracy. Reliable deployment requires robust data infrastructure, calibrated uncertainty, interpretable explanations, fair performance across contexts, and governance that aligns with public safety and sustainability goals. Aggregate morphology and mesostructural features provide physically meaningful inputs, but they must be captured with documented protocols and connected to mechanical tests that reflect field conditions. Model architectures should be designed for multimodal integration, distribution shift, and deployment constraints. Institutional practices such as model cards, datasheets, continuous monitoring, and human-in-the-loop review are essential. If these elements are treated as afterthoughts, deep learning risk models may become brittle black boxes that obscure physical mechanisms and undermine trust. If they are treated as core design requirements, deep learning can become a constructive part of materials informatics for asphalt pavements, supporting better mixture design, more efficient maintenance, and more sustainable infrastructure.

References

1. Roberts, F. L., Kandhal, P. S., Brown, E. R., Lee, D.-Y., & Kennedy, T. W. (1996). Hot mix asphalt materials, mixture design, and construction (2nd ed.). National Asphalt Pavement Association Research and Education Foundation.
2. Masad, E., Muhunthan, B., Shashidhar, N., & Harman, T. (1999). Internal structure characterization of asphalt concrete using image analysis. *Journal of Computing in Civil Engineering*, 13(2), 88–95.

3. Masad, E., & Buttar, W. G. (2004). Unified imaging approach for measuring aggregate angularity and texture. *Computer-Aided Civil and Infrastructure Engineering*, 19(4), 273–280.
4. Chandan, C., Sivakumar, K., Masad, E., & Fletcher, T. (2004). Application of imaging techniques to geometry analysis of aggregate particles. *Journal of Computing in Civil Engineering*, 18(1), 75–82.
5. Anochie-Boateng, J. K., Komba, J. J., & Mvelase, G. M. (2013). Three-dimensional laser scanning technique to quantify aggregate and ballast shape properties. *Construction and Building Materials*, 43, 456–465.
6. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
7. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097–1105.
8. L.W.Shi, Z.Y.Hu, H.M.Huang, et al. Effect of coarse aggregate morphology on asphalt mixture meso-structure using image recognition technology, *Measurement*, 2026, 290: 122757.
9. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
10. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.
11. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
12. Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. *Proceedings of the 33rd International Conference on Machine Learning*, 48, 1050–1059.
13. Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. *Proceedings of the 34th International Conference on Machine Learning*, 70, 1321–1330.
14. Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. *Advances in Neural Information Processing Systems*, 28, 2503–2511.
15. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229.
16. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, 64(12), 86–92.
17. Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A., Maharaj, T., Sherwin, E. D., Mukkavilli, S. K., Kording, K. P., Gomes, C. P., Ng, A. Y., Hassabis, D., Platt, J. C., ... & Bengio, Y. (2022). Tackling climate change with machine learning. *ACM Computing Surveys*, 55(2), 1–96.

18. Cha, Y.-J., Choi, W., & Büyüköztürk, O. (2017). Deep learning-based crack damage detection using convolutional neural networks. *Computer-Aided Civil and Infrastructure Engineering*, 32(5), 361–378.
19. Zhang, L., Yang, F., Zhang, Y. D., & Zhu, Y. J. (2016). Road crack detection using deep convolutional neural network. *2016 IEEE International Conference on Image Processing*, 3708–3712.
20. Tashman, L., Masad, E., D'Angelo, J., Bukowski, J., & Harman, T. (2003). X-ray tomography to characterize air void distribution in superpave gyratory compacted specimens. *International Journal of Pavement Engineering*, 4(1), 19–28.
21. Aschenbrener, T. (1995). Evaluation of Hamburg wheel-tracking device to predict moisture damage in hot-mix asphalt. *Transportation Research Record*, 1492, 193–201.
22. Federal Highway Administration. (2015). *Towards sustainable pavement systems: A reference document (FHWA-HIF-15-002)*. U.S. Department of Transportation.