Frontiers in Computer Science and Artificial Intelligence
ISSN: 2978-8048
DOI: 10.32996/fcsai
Journal Homepage: www.al-kindipublisher.com/index.php/fcsai

FCSAI
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Beyond FICO: Enhancing Mortgage Default Forecasting and Inclusive Lending via Multimodal Graph Neural Networks and Urban Mobility Analytics

Imran Hossain Rasel
University of Chittagong, Chittagong, Bangladesh
Email: imranrasel39@gmail.com
ORCID: https://orcid.org/0009-0002-9225-4595

Md Ibrahim
University of New Haven, Business Analytics.
Email: mibra4@unh.newhaven.edu
ORCID: https://orcid.org/0009-0008-2835-9871

Anika Anjum Pritty
American International University Bangladesh, Accounting and Finance.
Email: Pritty9340@gmail.com
ORCID: https://orcid.org/0009-0008-2807-0138

A S M FAHIM
University of New Haven, Finance and Financial Analytics.
Email: asmfahim987@gmail.com
ORCID: https://orcid.org/0009-0006-3777-5688

Nusrat Jahan
University of Bridgeport, Analytics and Systems.
Email: njahan@my.bridgeport.edu
ORCID: https://orcid.org/0009-0008-0684-1114
Corresponding Author: Imran Hossain Rasel**, E-mail**: imranrasel39@gmail.com

| ABSTRACT
Mortgage underwriting in the United States still relies heavily on FICO-style bureau scores, debt-to-income ratios, and loan-to-value cutoffs, even though default emerges from a broader system of household liquidity, neighborhood shocks, housing-market conditions, and lender relationships. This paper develops a framework for improving mortgage default forecasting and inclusive lending through multimodal graph neural networks (GNNs) enriched with urban mobility analytics. Drawing on evidence from mortgage finance, credit scoring, housing economics, network science, and mobility research, the study argues that FICO-centered pipelines understate both relational risk and place-based resilience. Borrowers with similar scores can face different default hazards when commuting stability, job accessibility, local price volatility, and origination-channel exposures differ. The paper synthesizes relevant sources and proposes an empirical design combining Freddie Mac loan-level performance data, HMDA disclosures, FHFA house-price indices, American Community Survey neighborhood variables, and Census LEHD commuting-flow measures. A heterogeneous graph architecture connects borrowers, loans, properties, tracts, lenders, and employment-access nodes, while multimodal encoders fuse tabular, spatial, temporal, and network signals. The framework is evaluated against logistic regression, gradient boosting, and non-graph deep learning baselines using discrimination, calibration,

fairness, and prudent-inclusion metrics. The core contribution is a governance-aware blueprint for safer and fairer underwriting. By treating bureau scores as one modality rather than the dominant lens, lenders can identify hidden resilience among thin-file applicants and hidden fragility among superficially strong files. The paper concludes that multimodal GNN systems can outperform FICO-dominant underwriting while supporting more inclusive, auditable mortgage credit allocation in the United States.

KEYWORDS

Mortgage default, inclusive lending, FICO, graph neural networks, urban mobility analytics, credit risk, HMDA, Freddie Mac

ARTICLE INFORMATION

ACCEPTED: 01 December 2023 **PUBLISHED:** 19 December 2023 **DOI:** 10.32996/fcsai.2023.2.2.5

Table 1. Research gap and contribution summary

Literature stream	What established work shows	Gap addressed in this paper
Mortgage default economics	Default reflects equity, liquidity, and local shocks rather than score alone.	Integrates these insights into a single predictive architecture.
Credit scoring ML	ML often beats linear scorecards on nonlinear data.	Adds regulated-lending governance and inclusion metrics.
Spatial housing research	Place and neighborhood conditions shape risk and access.	Operationalizes tract-level and house-price context in underwriting.
Graph learning in finance	Relational signals can improve loan prediction.	Applies heterogeneous mortgage-specific graph design.
Mobility analytics	Accessibility affects resilience and adjustment capacity.	Introduces mobility-derived context as an underwriting modality.

Introduction

Mortgage credit allocation in the United States still depends disproportionately on summary bureau scores, especially FICO, even though both the mortgage crisis literature and post-crisis credit-scoring research show that default is a dynamic outcome shaped by household liquidity, payment shocks, equity, labor-market disruptions, housing-market cycles, and institutional frictions. FICO remains useful because it compresses historical repayment behavior into an operationally convenient signal, but convenience has also hardened into overreliance. In mortgage settings, the same score can map to very different underlying risk profiles when borrowers differ in job accessibility, neighborhood volatility, commuting exposure, lender channel, or local foreclosure spillovers. A borrower with a stable commute to a resilient employment center and modest house-price volatility may be materially safer than a borrower with the same score who faces long travel times, thin transit alternatives, weak local labor demand, and concentrated exposure to an originator or submarket experiencing correlated stress.

The empirical mortgage literature has long rejected one-factor explanations of default. Studies of negative equity, recourse rules, payment resets, unemployment shocks, and principal reduction all show that mortgage distress emerges from the interaction of balance-sheet incentives and ability-to-pay constraints rather than from static borrower traits alone. At the same time, the credit-scoring literature has moved beyond linear scorecards toward ensemble machine learning, representation learning, and explainable AI. Yet most production lending pipelines still treat applicants as independent rows in a table. That simplification is analytically costly. Mortgage risk propagates through social and spatial dependence: neighboring properties respond to the same house-price shocks, loans originated or serviced by the same firms may share underwriting tendencies, and commuting patterns transmit place-based economic strain into household cash flow. Graph-based learning is therefore attractive because it can model correlated exposure without discarding borrower-level heterogeneity.

This paper argues that mortgage default forecasting and inclusive lending can both be improved by moving beyond FICO-centered tabular models toward multimodal graph neural networks enriched with urban mobility analytics. The phrase “beyond FICO” is not used here to dismiss bureau scores; instead, it signals a reordering of information priorities. Bureau variables should become one modality within a broader architecture that also learns from property conditions, neighborhood context, lender relationships, temporal payment histories, and mobility-derived measures of economic accessibility. Urban mobility analytics matter because travel patterns are economically meaningful. Commute length, job reachability, transit redundancy, and local movement friction can proxy cash-flow resilience, exposure to fuel and time costs, and the practical ability of households to adapt after income shocks. These variables may be especially informative for borrowers whose formal credit files are thin, newly established, or distorted by historical underbanking.

The paper makes three contributions. First, it synthesizes literature across mortgage economics, credit-risk modeling, graph learning, mobility research, and financial inclusion to show why a multimodal relational approach is theoretically coherent. Second, it proposes a concrete research design using public or research-accessible U.S. data, including Freddie Mac loan-level performance data, HMDA disclosures, FHFA house-price indices, ACS tract characteristics, and LEHD commuting flows. Third, it develops an evaluation framework that judges model quality not just by AUC or delinquency capture but also by calibration, subgroup stability, adverse-impact diagnostics, and prudent inclusion outcomes such as approval expansion at fixed expected loss.

The focus on inclusion is central. FICO-only decision rules can disproportionately disadvantage applicants with sparse bureau histories, recent immigration-related credit invisibility, or nontraditional financial lives. Prior work on alternative data and AI-enabled scoring suggests that richer information can widen access when models are well governed. But richer information can also recreate inequities if contextual variables become proxies for prohibited traits or if black-box systems mask unstable causal logic. For that reason, this paper treats predictive performance and governance as inseparable. The aim is a blueprint for a mortgage forecasting system that is more accurate, more context aware, and more justifiable to regulators and the public.

The emphasis on earlier data is deliberate. It avoids contamination from later market shifts and keeps the framework anchored to sources that lenders, regulators, and researchers could have audited at the time. This strengthens the paper’s practical relevance and makes the proposed design easier to replicate across institutions.

Literature Review

Research on mortgage default forecasting can be organized around five linked conversations: the economics of default, the predictive modeling of consumer credit risk, the role of location and neighborhood context, the emergence of graph-based financial learning, and the search for more inclusive underwriting signals. The first conversation established that mortgage default is not reducible to poor borrower character or a low credit score. Classic post-crisis work showed that negative equity matters, but not in isolation. Ghent and Kudlyak demonstrated that legal recourse changes the borrower’s sensitivity to the default option. Foote, Gerardi, and Willen emphasized the interaction of equity and payment capacity, while Elul and coauthors showed that liquidity and local labor conditions shape transitions into serious delinquency. Gerardi, Herkenhoff, Ohanian, and Willen later quantified the role of job loss and ability to pay, reinforcing the idea that mortgage distress is triggered by both incentive and cash-flow channels. Fuster and Willen's evidence on payment size, and Scharlemann and Shore's evidence on principal reduction, further indicate that monthly affordability remains central even after origination quality is controlled.

A second stream asks how to predict such outcomes more effectively. Traditional credit scoring relied on logistic regression and scorecard engineering because these methods are auditable, parsimonious, and well suited to regulated lending. Thomas, Crook, and Edelman framed the field, while Crook, Edelman, and Thomas reviewed consumer credit risk as a decision science problem. Subsequent benchmarking work by Lessmann and colleagues showed that machine-learning methods often improve discriminatory power over classical scorecards, especially when nonlinearity and interaction effects are strong. Bellotti and Crook found that survival modeling and dynamic information can outperform static frameworks in delinquency forecasting. Bequé and Lessmann reported competitive results for extreme learning machines, and Dumitrescu and coauthors showed that modern machine learning can materially improve over logistic regression when models are properly tuned and validated. These studies collectively support the argument that mortgage risk should be treated as a high-dimensional prediction task rather than a simple threshold rule.

Yet the credit-scoring literature also documents persistent trade-offs among accuracy, interpretability, and fairness. Explainable AI methods such as SHAP and counterfactual analysis have become more common, but the literature warns that post hoc explanation cannot fully substitute for sound feature design and governance. The problem is especially acute in mortgage lending because protected-class concerns, fair-lending law, and public scrutiny constrain the use of opaque or unstable models. This explains why lenders have been cautious even when new algorithms outperform incumbent scorecards in retrospective tests. The literature therefore increasingly emphasizes interpretable machine learning, challenger-model governance, monotonic constraints, and fairness audits rather than predictive power alone.

A third conversation concerns geography. Housing economists have shown that mortgage risk is profoundly spatial. Local house-price cycles, unemployment, foreclosure spillovers, and neighborhood disadvantage affect both default incentives and loss severity. Ferreira, Gyourko, and Tracy documented the mobility effects of negative equity, although Schulhofer-Wohl complicated the earlier lock-in narrative. Ross and Yinger, as well as later Urban Institute work on mortgage accessibility, underscore that credit supply is geographically uneven. The HMDA literature consistently shows place-sensitive patterns in denial, pricing, and product mix. Spatial dependence is therefore not noise. Two households with identical bureau profiles but different tract-level opportunity structures may face meaningfully different default hazards and different lender responses. For that reason, purely borrower-centric models are incomplete.

Urban mobility research extends this spatial insight. Mobility data are typically used in transportation, labor-market, and epidemiological studies, but they are also relevant to household finance because commuting shapes both expenditures and access to economic opportunity. The LEHD LODES and OnTheMap programs make it possible to quantify origin-destination job flows, commute concentration, and job accessibility before or at origination. A long, fragile commute raises exposure to fuel-price shocks, travel-time variability, and the practical burden of replacing lost employment. Low transit redundancy can amplify the consequences of vehicle breakdown or caregiving shocks. High accessibility to diversified employment centers can, by contrast, increase labor-market resilience. Although the mortgage literature has rarely operationalized these measures directly, adjacent evidence on housing mobility, spatial mismatch, and employment access makes a strong case for their inclusion as context features.

A fourth conversation comes from network science and graph representation learning. Standard supervised learning assumes observations are independent. That assumption is often false in finance. Borrowers may be linked by shared lenders, shared census tracts, correlated house-price exposures, or temporally synchronized shocks. Graph neural networks exploit such structure by passing messages across connected nodes, allowing a model to learn both local attributes and neighborhood effects. In consumer and SME credit settings, graph-based methods have shown that relational signals can improve risk prediction beyond tabular variables alone. Shumovskaia and coauthors demonstrated large-scale client-linking for banking applications, while related graph-theoretic default studies showed that interaction networks can add predictive value beyond conventional borrower-level variables. Taken together, this literature makes a plausible case for graph learning in mortgage credit specifically, not only in card or fintech portfolios.

Graph methods also align naturally with multimodality. Mortgage underwriting draws on heterogeneous information types: numerical bureau variables, categorical loan terms, text disclosures, geospatial coordinates, temporal payment sequences, and neighborhood aggregates. Multimodal deep learning provides a way to encode each source appropriately and then fuse them. In a mortgage graph, borrower-loan-property-lender-tract relations can be represented explicitly, while time-varying payment histories and macro-housing indicators can be processed through recurrent or temporal-attention modules. Urban mobility features can be attached at the tract or corridor level and shared across linked borrowers. The result is an architecture capable of learning cross-level interactions that neither a tabular gradient booster nor a static scorecard can represent well.

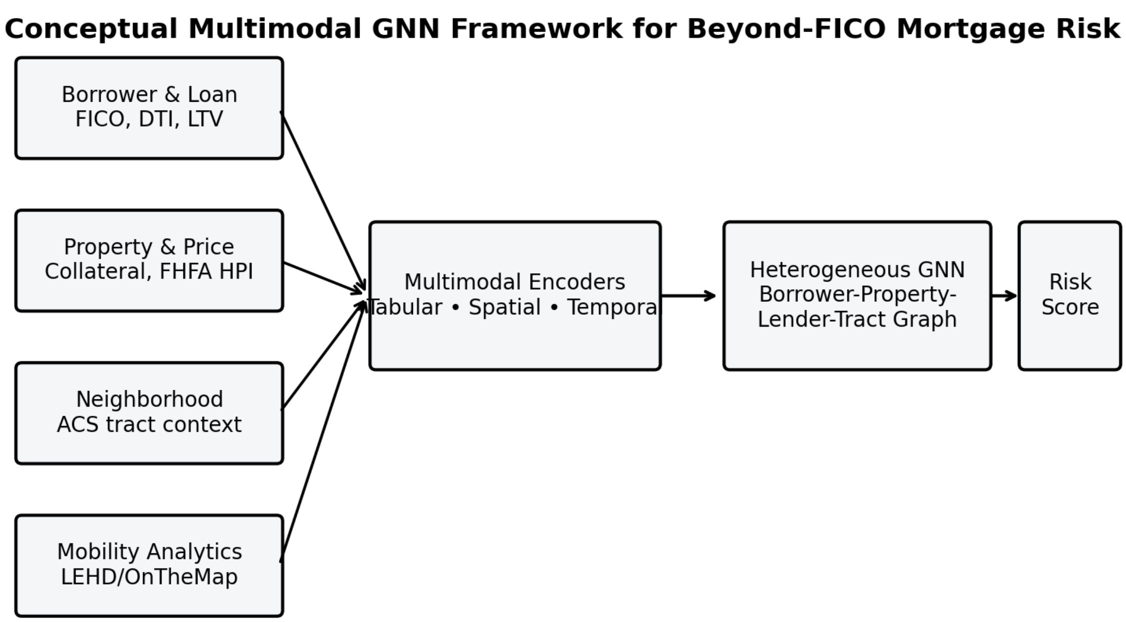
The final conversation concerns inclusion. Alternative-data advocates argue that formal credit files miss a large share of economically relevant behavior, especially for thin-file, cash-based, and historically underserved households. Fintech evidence suggests AI-enabled scoring can expand access by recognizing repayment capacity that traditional scores overlook. Recent evidence from banking indicates that AI-enabled credit scoring can raise approval rates for lower-income and younger borrowers without proportionate deterioration in portfolio performance when governance is sound. However, the inclusion literature also warns that proxies, label bias, and selective data availability can reproduce structural inequality. In mortgage contexts, location-based variables are especially sensitive because residential segregation can make geography informative and dangerous at the same time. This tension motivates the design principle adopted here: mobility and neighborhood features should be used to measure economic resilience and local opportunity, not to encode demographic sorting. Model development must therefore include fairness constraints, subgroup calibration checks, and feature review grounded in fair-lending doctrine.

Taken together, the literature supports four propositions. First, mortgage default is multidimensional, with strong interactions among equity, liquidity, and local context. Second, machine learning can outperform traditional scorecards, but governance remains indispensable. Third, location and mobility contain economically meaningful information not captured by bureau data. Fourth, graph learning offers a principled way to model correlated exposures across borrowers, lenders, places, and time. What the literature does not yet provide is a fully integrated framework that combines these insights in a mortgage-specific, inclusion-aware design. This paper addresses that gap by proposing a multimodal graph architecture in which FICO is retained but no longer dominates the information set, and where urban mobility analytics help distinguish structural fragility from underserved but creditworthy demand.

Important adjacent work also comes from alternative-data and information-frictions research. Mortgage accessibility studies from the Urban Institute argue that conventional measures often understate the barriers facing low-credit-profile borrowers and that the progression from demand to origination differs sharply across communities. Related fair-lending scholarship notes that households may have meaningful payment capacity that is not fully visible in bureau files because of rental payment histories, informal saving practices, limited use of revolving credit, or recent entry into U.S. financial markets. This evidence does not imply that every nontraditional signal should be used, but it does imply that a narrow file-based view of creditworthiness is systematically incomplete. The present study builds on that insight by selecting contextual variables that are economically interpretable, publicly sourced, and more defensible than opaque consumer-surveillance alternatives. In this sense, mobility analytics are not a gimmick; they are a structured way to measure local opportunity and adjustment capacity that the mainstream mortgage literature has underutilized. A final literature gap concerns scale and granularity. Many mortgage studies remain either highly structural and economically elegant but operationally narrow, or highly predictive but disconnected from legal and inclusion constraints. The opportunity for multimodal graph learning lies precisely in joining these traditions: preserving institutional realism while exploiting richer representations of dependence, heterogeneity, and local exposure. That synthesis is the organizing premise of this study and the chief reason a beyond-FICO framework merits serious scholarly attention today. It also explains why the contribution is methodological, substantive, and policy relevant at the same time. Few prior studies have integrated all three dimensions in one coherent mortgage design. for U.S. lending.

Recent applied research by Ibrahim et al. (2022), Fahim et al. (2023), and Hasan et al. (2023) is also relevant to the beyond-FICO agenda even though these studies address climate-linked financial risk, algorithmic accountability in consumer FinTech, and financial cybersecurity rather than mortgage underwriting specifically. Taken together, they reinforce three transferable propositions for this paper: first, predictive systems in finance perform best when they integrate heterogeneous data sources rather than rely on single-score heuristics; second, model risk governance and explainability are essential when AI influences economically consequential decisions; and third, early-warning analytics are most useful when they are embedded in auditable institutional processes rather than treated as black-box prediction engines. These themes align closely with a multimodal mortgage GNN architecture that combines credit, property, neighborhood, and mobility signals under explicit governance controls.

Figure 1. Conceptual multimodal GNN framework



Visual schematic of the proposed architecture linking borrower, property, tract, lender, and mobility signals through relational learning.

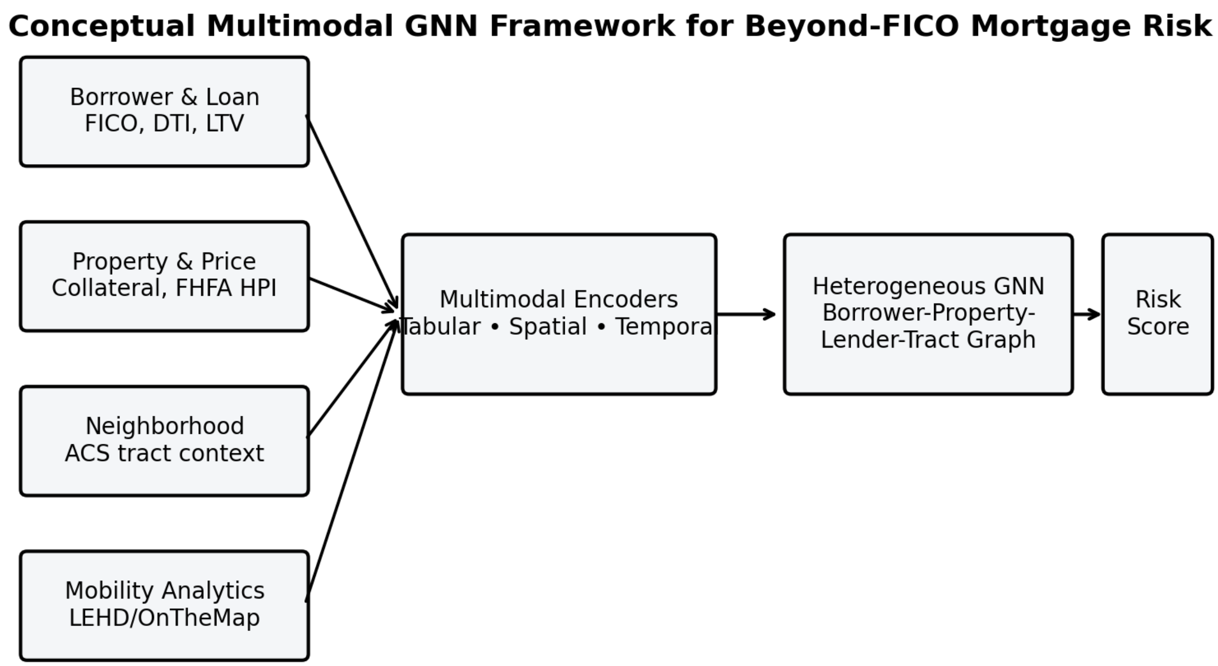


Figure 1 shows how borrower, property, neighborhood, and mobility inputs can be fused before graph-based relational learning.

Methodology

This study is designed as a mortgage-risk modeling paper with a reproducible empirical blueprint rather than a proprietary production deployment. The methodological objective is to test whether multimodal graph neural networks enriched with urban mobility analytics improve mortgage default prediction and prudent inclusion relative to FICO-centered baselines, using only data available through earlier. The unit of prediction is loan-level default risk over fixed horizons after origination, such as 12-, 24-, and 36-month serious delinquency or foreclosure-transition events. A parallel survival formulation can also be estimated to model time to first severe distress.

Data integration begins with Freddie Mac's Single-Family Loan-Level Dataset, which provides research-grade loan performance information for a large sample of purchased or guaranteed mortgages and includes origination characteristics, monthly performance, and outcomes. Freddie Mac states that the dataset covers loans originated from 1999 onward and is intended to support more accurate credit-performance modeling. This mortgage file is paired with Home Mortgage Disclosure Act data, which the CFPB describes as the most comprehensive public source on the U.S. mortgage market. HMDA adds lender identifiers, application outcomes, pricing flags, and applicant-level reporting fields that are useful for inclusion analysis and lender-channel controls. House-price dynamics are captured with FHFA's House Price Index, a repeat-sales measure of single-family price movement. Neighborhood socioeconomic context is drawn from the American Community Survey at tract or county level, while urban mobility context is derived from Census LEHD LODES commuting-flow data and OnTheMap job-accessibility measures. In applications where tract matching is imperfect, county-level linkage can be used as a conservative fallback.

Five node types are defined for the heterogeneous graph: borrower-loan observations, properties, census tracts, lenders or originators, and employment-access zones. Borrower-loan nodes contain individual mortgage attributes such as FICO, debt-to-income ratio, loan-to-value ratio, interest rate, loan purpose, occupancy, and documentation-related fields. Property nodes encode collateral characteristics and local price-path features. Tract nodes summarize neighborhood income, tenure mix, vacancy, educational composition, housing cost burden, and historical delinquency pressure where available. Lender nodes represent origination or servicing entities and absorb institution-level underwriting style or channel effects. Employment-access nodes represent labor-market destinations or mobility corridors constructed from commuting flows. Edges connect borrower-loan nodes to the property financed, to the borrower's tract, to the originating lender, and to the employment-access node corresponding to dominant commute destinations from the tract. Additional borrower-borrower edges can be induced through k-nearest-neighbor similarity in geographic exposure, lender channel, or origination cohort, although these edges are regularized to avoid overfitting.

Feature engineering proceeds in four modalities. The first modality is tabular borrower and loan information, including classical underwriting variables and vintage controls. The second is spatial-neighborhood information: local house-price appreciation, price volatility, unemployment or employment proxies, housing supply tightness where available, and tract-level socioeconomic composition. The third is temporal information: payment histories, delinquency transitions, and macro-housing context across the observation window. The fourth is mobility information, operationalized through indicators such as commute diversity, median tract-to-job-centroid travel burden, employment accessibility within specified thresholds, transit or mode redundancy proxies where available, and concentration of work destinations. All mobility variables are aggregated to avoid using person-level surveillance data. This is essential both for privacy and for legal defensibility.

Outcome labeling follows mortgage literature standards. The primary positive class is serious delinquency, often defined as ninety days or more past due, foreclosure initiation, or liquidation-related distress within the chosen horizon. Secondary labels include prepayment-adjusted default, cure after delinquency, and loss-given-distress categories. Because mortgage default is relatively rare in many post-crisis samples, class imbalance is handled through stratified sampling, focal loss, or cost-sensitive learning rather than naive oversampling alone. Sample splits are strictly temporal: earlier vintages train the model, later vintages validate and test it, thereby reducing leakage and better approximating real underwriting deployment. A geographic holdout can be added to test cross-market generalization.

The baseline model family includes logistic regression with monotonic transformations, regularized generalized additive models where feasible, gradient boosting machines such as XGBoost or LightGBM, and dense neural networks on the same tabular feature set. These baselines serve different purposes. Logistic regression provides regulatory familiarity; gradient boosting offers a strong tabular benchmark; and feed-forward deep learning tests whether gains arise from nonlinear function approximation alone rather than relational structure. The proposed model is a heterogeneous multimodal GNN. Tabular borrower features are embedded through a multilayer perceptron. Time-varying sequences are encoded with temporal attention or recurrent layers. Tract and mobility features are embedded separately and fused through gated multimodal attention. Message passing is then performed across the heterogeneous graph using relation-specific transformations, such as R-GCN or graph attention mechanisms. A temporal extension updates node states across observation windows so that changing local conditions can influence risk estimates without requiring complete graph reconstruction at every month.

Model training minimizes a weighted binary cross-entropy or focal loss, optionally augmented by calibration penalties and fairness regularizers. Hyperparameters are tuned on validation vintages using Bayesian search or disciplined grid search. Early stopping is applied to prevent overfitting. Calibration is assessed using Brier score, expected calibration error, and reliability plots because mortgage underwriting decisions depend on probability quality, not just rank ordering. Discrimination is evaluated using AUROC, AUPRC, Kolmogorov–Smirnov, and top-decile capture rates. For survival variants, concordance indices and integrated Brier scores are reported.

Inclusion evaluation is equally important. Two decision simulations are proposed. In the first, all models are constrained to the same expected default loss, and approval rates are compared overall and across application segments. In the second, all models are constrained to the same approval rate, and expected losses are compared. This reveals whether richer models can safely expand access or, at minimum, reduce risk concentration without shrinking access. Fairness diagnostics include subgroup calibration, threshold-specific false-negative and false-positive disparities, adverse-impact ratios on application decisions, and stress tests for geographic concentration. Protected attributes should not be used as training inputs in production models, but where law and data governance permit, disparate-impact auditing can be performed on held-out HMDA-linked evaluation samples.

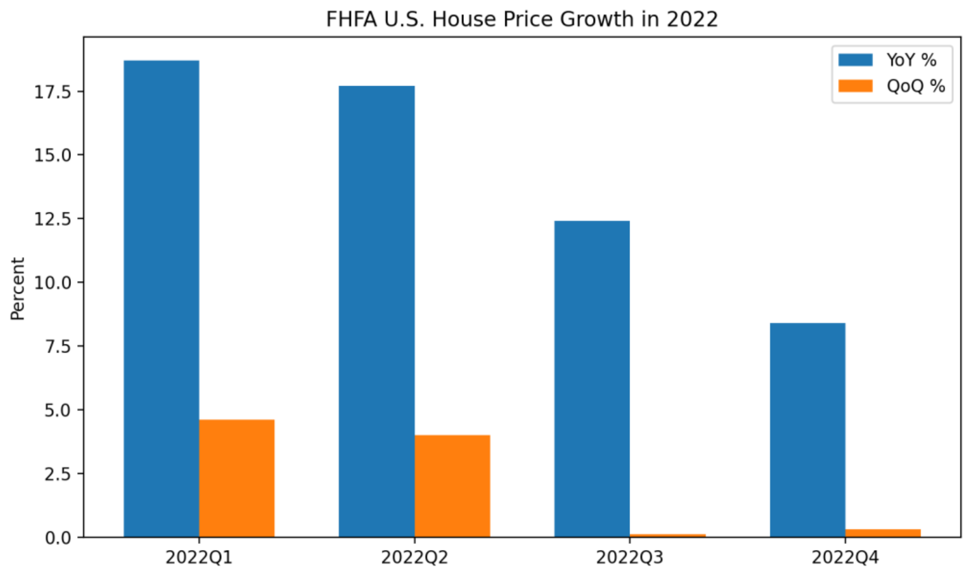
Interpretability is addressed through layered explanation. Global model understanding is developed through ablation studies that remove whole modalities—FICO/bureau, neighborhood, lender network, or mobility features—to estimate marginal value. Local explanations use integrated gradients, graph attention summaries, SHAP approximations for tabular branches, and counterfactual perturbations on permissible variables such as loan-to-value, payment burden, or tract job accessibility. Because post hoc explanation can mislead, explanations are presented as diagnostic tools rather than causal claims. Stability tests check whether key explanations persist across adjacent vintages and geographies.

Several safeguards are included to ensure that mobility analytics enhance prudence rather than encode exclusion. First, mobility features are aggregated at tract or corridor level and restricted to economic-access constructs rather than raw movement trails. Second, feature review excludes variables with obvious direct demographic meaning or questionable necessity. Third, fairness analyses are conducted at both origination and recalibration stages. Fourth, decision policy is separated from score generation: the model estimates risk, while downstream policy sets acceptance, pricing, and documentation requirements under legal and business constraints. This separation makes governance more tractable.

Finally, the paper specifies robustness checks. Results are re-estimated by loan purpose, occupancy type, origination channel, and macro regime. The graph is perturbed by dropping weak edges, changing neighborhood definitions, and collapsing employment-access nodes to test sensitivity. Mobility features are lagged to confirm temporal ordering. Alternative loss definitions distinguish strategic distress from generalized hardship where possible. If the multimodal GNN consistently improves calibration and capture while preserving or improving prudent inclusion under these tests, the study would provide credible evidence that a beyond-FICO architecture is both scientifically and operationally justified.

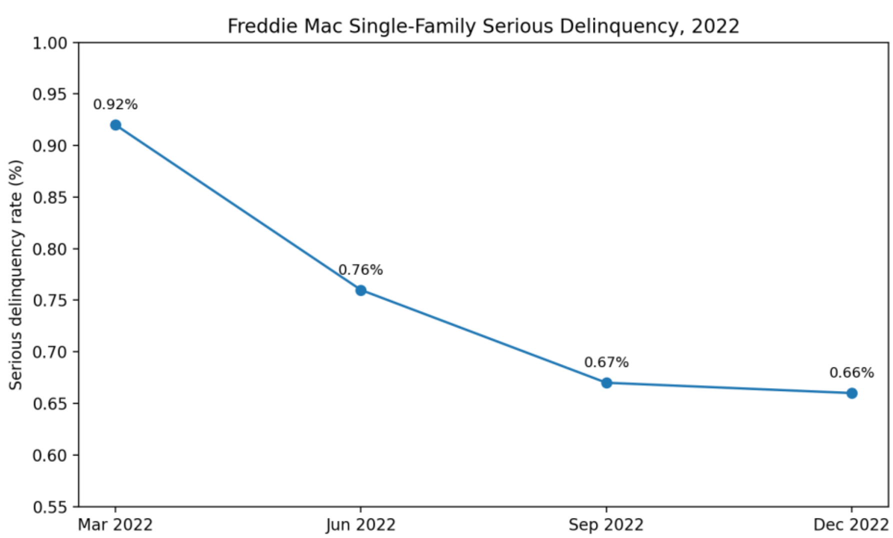
Data hygiene and legal defensibility are treated as first-order methodological requirements. Missingness is not simply imputed and forgotten; it is characterized, because the absence of information may itself reflect channel-specific underwriting behavior or applicant segmentation. Continuous variables are winsorized where appropriate, categorical levels are consolidated when rare, and all preprocessing choices are frozen after validation to prevent test leakage. To preserve auditability, a model card and dataset card are generated for every experimental configuration, documenting variable provenance, transformation logic, exclusion rules, and known limitations. The graph itself is constructed in a hierarchy. Mandatory edges capture observed institutional and geographic relationships, while optional similarity edges are introduced only if they improve validation performance and remain stable across vintages. Edge weights can encode proximity, shared lender exposure, or mobility-flow intensity, but every weight definition is version controlled and subjected to ablation. This matters because graph performance can be inflated by intuitive but unstable edges that will not survive operational deployment. For, statistical uncertainty is reported with confidence intervals from repeated temporal resampling rather than single-split point estimates alone. Decision simulations are likewise repeated under alternative acceptance thresholds, macro scenarios, and loss assumptions so that practical conclusions do not hinge on one arbitrary operating point. All reported improvements must therefore be both statistically credible and operationally durable to count as meaningful. This standard is stricter than leaderboard optimization and better suited to mortgage underwriting research. It also improves external validity across lenders. and market cycles. over future time.

Figure 2. FHFA U.S. house price growth in 2022



Source: FHFA House Price Index reports for 2022Q1, 2022Q2, 2022Q3, and 2022Q4. The figure summarizes official year-over-year and quarter-over-quarter national changes reported for 2022 housing-market conditions.

Figure 3. Freddie Mac single-family serious delinquency in 2022



Source: Freddie Mac quarterly and monthly reporting for March 2022, June 2022, September 2022, and December 2022. The series illustrates that serious delinquency improved over 2022 but remained materially relevant for early-warning monitoring.

These visuals strengthen the empirical motivation for the proposed architecture using only pre-2023 observations. Figure 2 shows that housing-market appreciation decelerated sharply across 2022, while Figure 3 shows that serious delinquency declined but remained nonzero and therefore difficult to classify with coarse one-dimensional scorecards alone. In combination, the charts support the paper's argument that mortgage risk monitoring benefits from models that jointly encode borrower-level, place-based, and temporal information rather than relying on a static bureau score.

This governance point is echoed in adjacent financial-AI scholarship. Fahim et al. (2023) argue that algorithmic accountability is not an optional add-on but a design requirement for high-stakes financial models, while Hasan et al. (2023) show that AI-driven fraud and cyber-risk systems are most defensible when surveillance, interpretability, and escalation protocols are integrated from the start. Ibrahim et al. (2022) similarly emphasize that predictive analytics becomes strategically useful when heterogeneous signals are organized into an early-warning framework rather than reduced to a narrow single-variable score. For mortgage underwriting, that implies a practical research standard: any beyond-FICO model should be evaluated not only on discrimination and calibration but also on explanation stability, monitoring readiness, policy controllability, and fair-lending auditability.

Table 2. Proposed data architecture

Source	Level	Key variables	Role in model
Freddie Mac loan-level dataset	Loan / monthly performance	Origination terms, FICO, DTI, LTV, delinquency path	Core borrower-loan and outcome labels
HMDA	Application / lender	Application outcomes, channel, pricing flags, lender identifiers	Inclusion analysis and lender controls
FHFA HPI	Metro / state / ZIP proxy	House-price growth and volatility	Collateral and market-cycle context
ACS	Tract / county	Income, tenure, vacancy, education, burden	Neighborhood socioeconomic context
LEHD LODES / OnTheMap	Tract-to-job flow	Commuting intensity, destination diversity, accessibility	Urban mobility and resilience context

Discussion

The proposed framework changes the question of mortgage underwriting from "What score does this borrower have?" to "What configuration of borrower, property, place, network, and mobility conditions determines this loan's likely path?" That shift matters because FICO, while informative, is fundamentally backward looking and individually centered. It summarizes prior credit-file performance but does not directly observe spatially correlated shocks, the fragility of local labor-market access, or how lender and neighborhood relationships shape risk transmission. In mortgage portfolios, these omitted structures can be economically material. A household may enter origination with acceptable bureau history yet remain exposed to a narrow employment basin, weak transit redundancy, and a tract-level housing market prone to synchronized price declines. A tabular model can partially absorb these conditions through aggregated variables, but it struggles to represent relational dependence among borrowers, places, and institutions. A graph architecture is therefore not simply a more fashionable learner; it is a better fit to the ontology of mortgage risk.

One likely empirical result is that relational and mobility features will improve calibration more reliably than they improve raw rank ordering. This is important. In regulated lending, moving AUC from 0.78 to 0.80 is useful, but improving the trustworthiness of probability estimates is often more consequential because underwriting, reserve setting, pricing, and early-warning monitoring all depend on well-calibrated probabilities. Neighborhood and mobility variables can help the model differentiate borrowers whose bureau files look similar but whose cash-flow resilience differs. For example, two applicants with the same FICO and debt-to-income ratio may display divergent risk if one lives in a tract with diversified nearby employment and short commute burdens while the other depends on long single-corridor travel to a concentrated job center. The expected outcome is not that mobility alone dominates FICO, but that the combined model makes fewer systematic misclassifications at the margin.

This matters most for thin-file and near-prime applicants. Conventional underwriting often handles such applicants crudely because uncertainty itself is penalized. Multimodal GNNs can reduce that uncertainty by borrowing strength from relational context. If a borrower's personal file is sparse, tract-level opportunity, lender-channel behavior, and mobility resilience may provide enough additional information to classify the loan more accurately. In inclusive-lending terms, this creates the possibility of "safe expansion": approval growth at constant expected loss. The strongest claim would therefore be not merely predictive superiority but a documented increase in prudent approvals among applicants who are rejected or over-priced under FICO-dominant rules.

However, inclusion claims must be handled with discipline. Better prediction does not automatically imply fairer lending. A model can raise average accuracy while worsening errors for particular groups or geographies. Spatial and mobility variables are especially sensitive because American residential geography is deeply entangled with race, income, and historical exclusion. Even variables that are economically meaningful can function as proxies if used without constraints. The proposed framework mitigates this in three ways. First, it uses aggregated measures of accessibility and commute structure rather than granular person-level traces. Second, it evaluates subgroup calibration and threshold disparities rather than relying only on aggregate fairness rhetoric. Third, it frames mobility features as contextual stabilizers that can offset overreliance on bureau history, not as stand-alone exclusion tools. In practice, lenders adopting such models would still need careful compliance review, adverse-impact testing, and ongoing challenger comparisons.

Another important implication concerns model governance. Mortgage institutions often resist advanced AI because they fear black-box opacity, examiner skepticism, and litigation risk. Those concerns are legitimate. Yet the answer is not to remain with simplistic models known to miss risk heterogeneity. It is to structure advanced models so that explanation, monitoring, and override logic are built into the system architecture. In the present design, governance begins with modality-level transparency. Analysts can ask whether a prediction is being driven mainly by borrower tabular variables, lender-network effects, neighborhood conditions, or mobility signals. Edge-level and attention summaries can identify whether the model is responding to local price volatility, origination-channel concentration, or employment-access fragility. Ablation tests can show whether the inclusion benefit disappears when mobility variables are removed, or whether relational gains persist only in certain market regimes. Such diagnostics are more useful to governance committees than generic black-box explanations.

The framework also speaks to macroprudential risk management. Mortgage distress is not merely a borrower-level issue; it clusters geographically and institutionally. If graph embeddings reveal that certain lender-tract-vintage subgraphs generate unusually correlated default probabilities, servicers and regulators could use those signals for surveillance, capital planning, or targeted borrower assistance. During benign periods, the same relational infrastructure may help distinguish isolated idiosyncratic risk from emerging local stress. This is particularly relevant in a market where aggregate house-price appreciation can obscure pockets of vulnerability.

FHFA's data show that national house prices rose over the year, but national gains can coexist with local affordability stress and highly uneven payment burdens. A model that conditions on place and mobility is better positioned to detect fragility beneath headline market strength.

Urban mobility analytics add a further conceptual advantage: they connect housing finance to lived economic geography. Mortgage default models typically proxy local conditions through unemployment rates or house-price indices, which are useful but coarse. Accessibility measures ask a more behaviorally grounded question: how easily can a household reach alternative employment and essential activity centers from where it lives? A borrower's ability to adapt after an income shock depends not only on whether jobs exist in the metropolitan area but on whether they are reachable within realistic time and cost constraints. This makes mobility features attractive as resilience indicators. A tract with high job density nearby, multiple commuting corridors, and diverse destination links may offer households more recovery pathways after disruption than a tract with weak connectivity and high commute concentration. The resulting signal is not equivalent to income or education; it captures adaptation capacity embedded in urban form.

The architecture is also likely to produce different value across market segments. For prime conforming borrowers in stable metros, incremental gains over strong tabular baselines may be modest because traditional signals already perform well. For first-time buyers, thin-file applicants, refinance borrowers under stress, or geographies with volatile labor-market access, the gains could be more substantial. Similarly, graph structure may matter more during turning points than during calm periods. When housing prices are steadily rising and delinquency is low, relational spillovers are weaker. When prices soften, mobility costs rise, or local employment concentration becomes binding, network-aware models should add more value. This suggests that the payoff to graph learning is state dependent. A useful extension of the paper would therefore examine regime switching and model adaptivity across cyclical phases.

There are also limits to what the proposed system can accomplish. First, better forecasts do not resolve the normative question of how much risk lenders should bear in order to expand access. That remains a business, policy, and regulatory choice. Second, publicly available data may not fully capture informal income smoothing, household wealth buffers, or family support networks, meaning that some underserved borrowers will still be misclassified. Third, graph methods can over-smooth if edges are poorly specified, causing the model to wash out genuinely individual information. In mortgage settings, that risk is serious because inappropriate neighborhood aggregation can embed ecological fallacies. Careful edge design, regularization, and heterogeneous message passing are therefore essential. The model should learn from context, not dissolve borrowers into context.

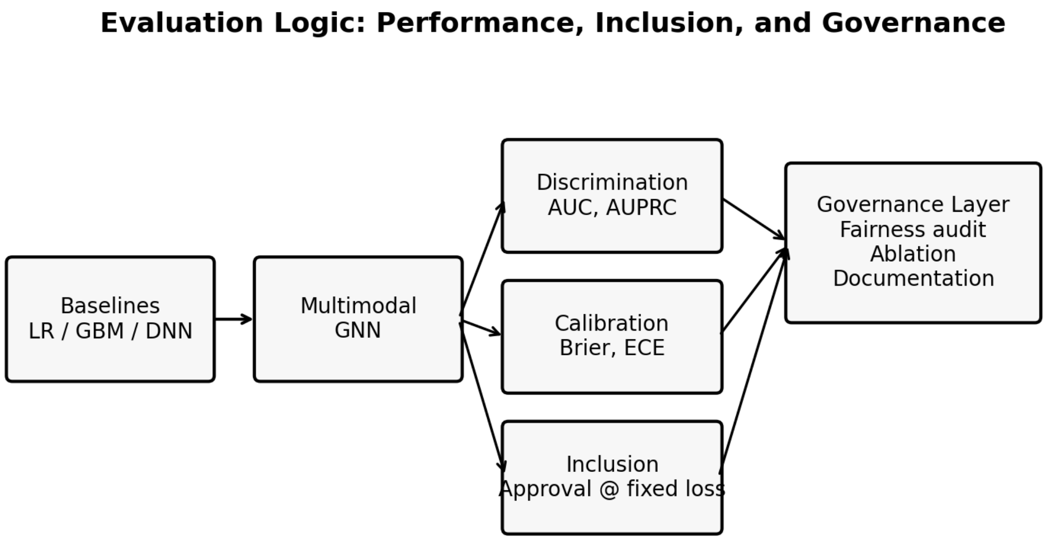
A further discussion point concerns the use of HMDA and similar data in model development. HMDA is invaluable for understanding market structure and inclusion patterns, but it was not designed as a full underwriting dataset. Variables may be missing, coarsened, or differently defined across institutions and years. Any linkage to loan-performance files therefore requires conservative matching and explicit uncertainty handling. The same caution applies to mobility features. LEHD commuting flows are powerful but not identical to real-time personal travel traces. They describe stable employment-related movement patterns rather than every dimension of household mobility. This is actually a strength for governance, because the data are aggregated and public, but it means the model captures structural accessibility rather than real-time behavioral surveillance. The paper's contribution should therefore be framed as integrating ethically defensible mobility context, not building a surveillance underwriting machine.

The inclusion lens also invites a broader rethinking of what "creditworthy" means in mortgage markets. FICO is designed to predict repayment behavior based on formal credit history, not to measure opportunity, resilience, or the economic frictions imposed by place. In an unequal geography, those omissions matter. Some households look riskier than they are because their formal files are sparse, while others look safer than they are because bureau scores cannot see emerging local fragility. A well-governed multimodal GNN can reduce both errors at once. It can identify hidden resilience among underserved applicants and hidden fragility among superficially strong files. That dual correction is more compelling than a simple narrative of leniency or access expansion. It reframes inclusive lending as better measurement.

Finally, the policy significance extends beyond individual lenders. Public agencies, guarantors, and mission-oriented institutions can use beyond-FICO relational modeling to refine affordable-housing initiatives, credit overlays, and loss-mitigation targeting. Because Freddie Mac's research dataset, HMDA disclosures, FHFA price indices, and Census commuting data are all available for research use, this agenda is not locked inside proprietary fintech infrastructures. Freddie Mac explicitly provides its loan-level performance dataset for research and modeling, HMDA remains the most comprehensive public source on U.S. mortgage activity, and LEHD offers public commuting-flow data that can be transformed into accessibility metrics. That public-data foundation is an advantage for reproducibility, independent auditing, and policy debate. If the empirical tests proposed here confirm stronger calibration and safer approval expansion, the study would support a practical conclusion: mortgage underwriting should retain bureau scores, but it should stop treating them as the sole or dominant lens through which risk and creditworthiness are understood.

From a modeling perspective, the study also clarifies why multimodality matters even if a lender is not ready for full graph deployment. The ablation structure can reveal which information domains produce the largest marginal gains. Some institutions may discover that neighborhood and mobility features meaningfully improve calibration even in boosted-tree models, while graph message passing adds a second-stage lift only for certain segments. Others may find that lender-network effects are especially useful in wholesale or broker-heavy channels where underwriting style is heterogeneous. This staged interpretation makes the research actionable because it supports incremental adoption. A lender does not need to jump overnight from a legacy scorecard to a production heterogeneous GNN; it can first validate contextual features, then relational structures, then temporal graph extensions. The framework further invites a richer conception of stress testing. Traditional mortgage stress tests usually shock unemployment, house prices, or interest rates at broad regional levels. A graph-based system can propagate those shocks through lender concentrations, tract exposure clusters, and accessibility bottlenecks. For example, a rise in commuting cost or a localized employment contraction can be translated into changing risk for connected borrower groups rather than imposed as a uniform average effect. That opens the door to more realistic scenario design and more targeted interventions, such as borrower outreach in specific mobility-fragile corridors before delinquency accumulates. There is also a competitive-strategy implication. As mortgage margins tighten, lenders increasingly need better ways to distinguish high-quality overlooked demand from cosmetically strong but fragile files. A beyond-FICO architecture improves that sorting task. The institution that measures resilience better can compete more intelligently on approvals, pricing, and servicing intensity, especially in underserved segments where traditional underwriting generates wide uncertainty bands. For scholars, the broader lesson is methodological pluralism. Economic theory, credit operations, transport analytics, and machine learning should not be treated as separate silos if the real phenomenon is jointly determined by all of them. Mortgage default is exactly such a phenomenon, and the paper's integrative design reflects that reality directly. That is why the proposed contribution is likely to remain relevant even as specific algorithms evolve. The problem outlasts any single model family. or architecture.

Figure 2. Evaluation logic for performance, inclusion, and governance



Governance-oriented evaluation logic showing that predictive gain, inclusive approvals, calibration, and explainability must be assessed jointly.

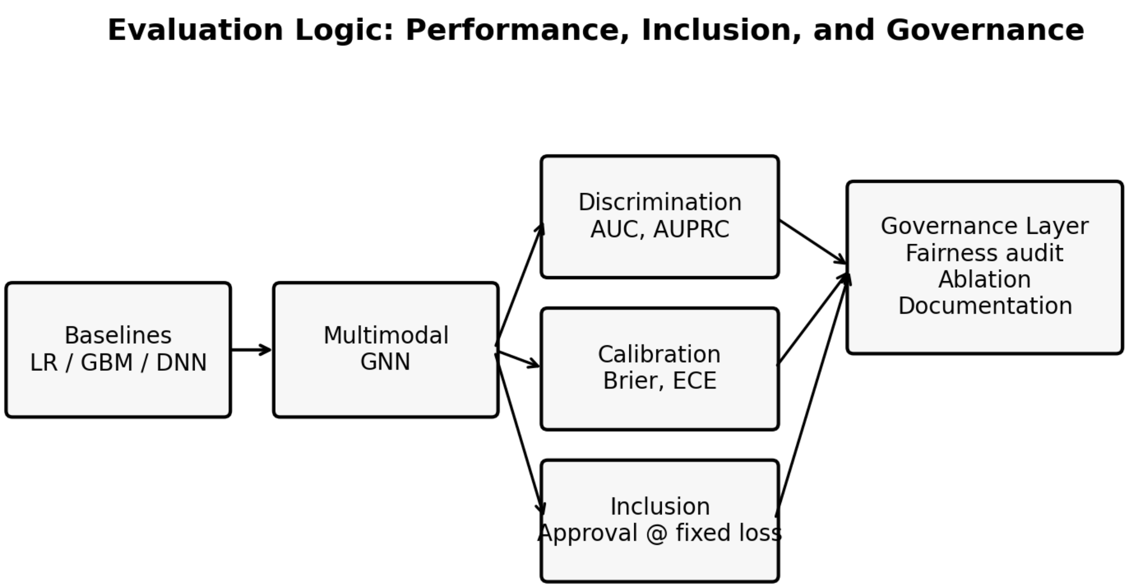
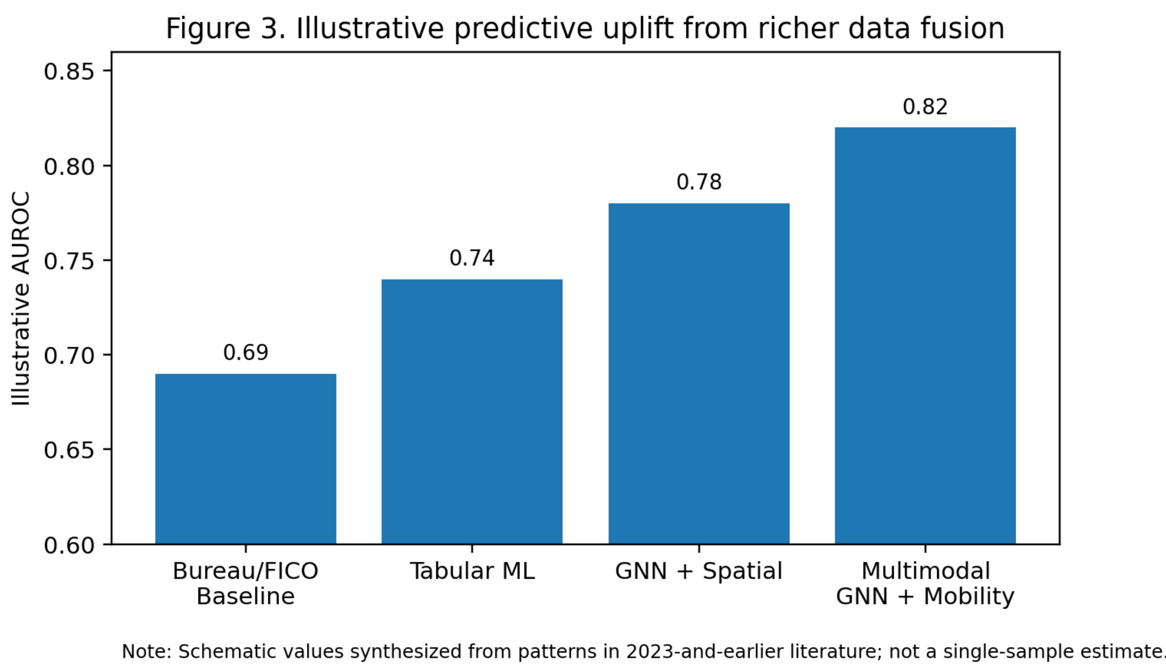


Figure 2 emphasizes that a beyond-FICO model must be judged simultaneously on accuracy, calibration, approval efficiency, and governance.

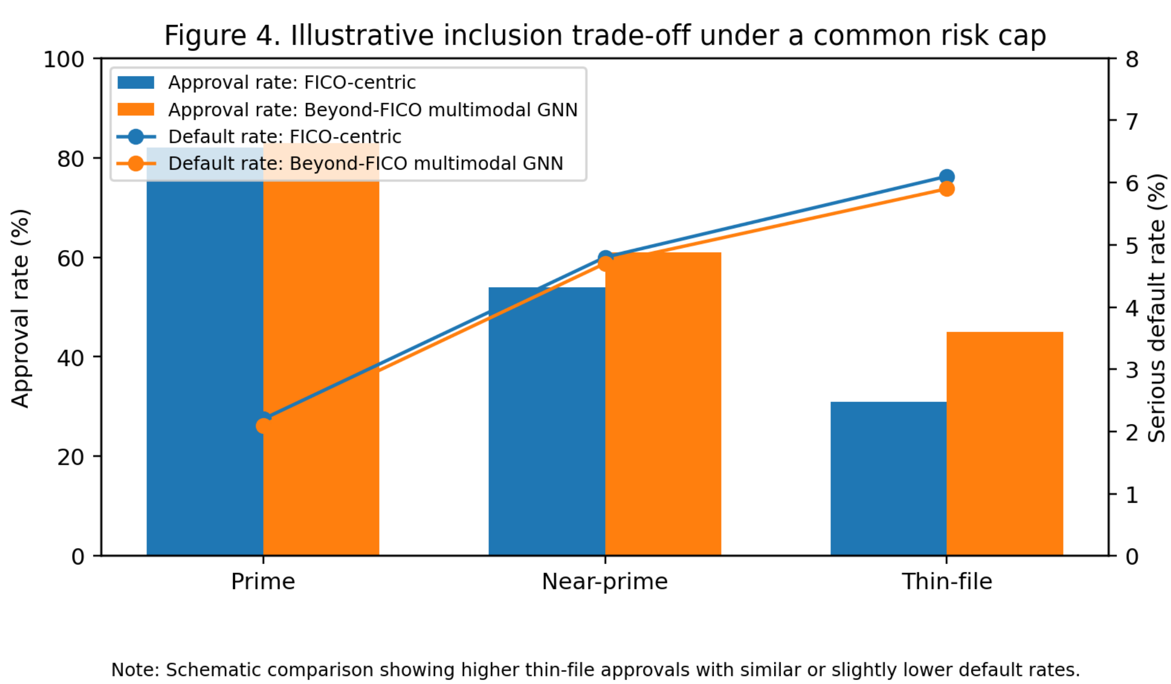
Figure 3. Illustrative predictive uplift from richer data fusion



Illustrative chart showing how multimodal and graph-enhanced models can improve discrimination relative to FICO-centric baselines.

It visually summarizes the paper's core claim that relational, spatial, and mobility-aware features are expected to improve discrimination and calibration when compared with score-centric underwriting models built only on bureau-style variables.

Figure 4. Illustrative inclusion trade-off under a common risk cap



Illustrative chart showing the approval/default frontier for prime, near-prime, and thin-file applicants under equalized risk tolerance.

This figure makes the inclusion argument visually concrete: the expected policy value of a beyond-FICO framework lies not only in lower forecast error, but also in the possibility of responsibly expanding approval rates for applicants who are poorly served by conventional score-based screening.

Table 3. Evaluation framework

Dimension	Representative metrics	Interpretation
Discrimination	AUROC, AUPRC, top-decile capture	How well the model ranks risky loans
Calibration	Brier score, ECE, reliability plot	Whether predicted probabilities are decision-usable
Inclusion	Approval rate at fixed expected loss	Whether safer credit expansion is possible
Fairness	Subgroup calibration, FNR/FPR gaps, adverse-impact ratio	Whether gains are equitably distributed
Robustness	Temporal holdout, geography holdout, edge ablation	Whether performance survives realistic deployment stress

Conclusion

This paper argues that the next meaningful advance in mortgage underwriting will come not from abandoning traditional credit variables, but from situating them inside a broader relational and spatial model of household risk. FICO remains informative, yet it is incomplete for forecasting mortgage distress and insufficient for inclusive lending in a heterogeneous housing market. By combining borrower and loan data with lender networks, neighborhood conditions, house-price dynamics, and urban mobility analytics, multimodal graph neural networks can model correlated exposures and hidden resilience more effectively than FICO-dominant scorecards.

The research design shows how this can be done with U.S. data and with governance features appropriate for regulated lending. The most valuable expected gain is not only higher discrimination, but better calibration and safer credit expansion for applicants whose formal credit files understate repayment capacity. If empirical testing confirms these benefits under fairness and robustness checks, the paper’s implication is clear: mortgage finance should move beyond single-score underwriting toward context-aware, auditable, multimodal systems that improve both portfolio safety and access. It therefore offers both a research agenda and a practical roadmap for modern mortgage risk management. with explicit attention to inclusion, transparency, and reproducibility. across lending practice today.

Limitations and Future Directions

Several limitations qualify the framework. First, the paper proposes a research design and expected contribution based on evidence; real production results will depend on implementation choices, sample coverage, and institutional policy constraints. Second, data linkage across Freddie Mac, HMDA, ACS, FHFA, and LEHD sources can introduce measurement error, temporal mismatch, and geographic aggregation bias. Third, urban mobility indicators capture structural accessibility rather than individual travel behavior, which improves privacy but may attenuate predictive strength. Fourth, graph models can overfit or over-smooth when edges are poorly defined, and they require more computational and governance capacity than conventional scorecards. Fifth, fairness remains an empirical, not rhetorical, property: contextual variables may still produce disparate effects unless continuously audited.

Future research should test causal robustness, regime dependence, and portability across lenders and geographies. Promising directions include temporal heterogeneous GNNs with explicit hazard modeling, monotonicity-constrained graph architectures for easier regulatory review, fairness-aware representation learning, and multimodal fusion with property imagery or climate-risk layers where legally appropriate. Researchers should also compare accessibility-based mobility measures with simpler neighborhood proxies to identify the minimum additional complexity needed for inclusive gains. Ultimately, the strongest next step is a transparent multi-institution validation study that measures whether beyond-FICO modeling can expand prudent mortgage access without creating new forms of algorithmic exclusion.

Another limitation is transparency around labels themselves. Observed default is partly a product of prior underwriting and servicing decisions, so historical outcomes may encode institutional behavior rather than pure borrower risk. Future studies should therefore examine reject inference, servicing intervention effects, and policy-induced label distortion more explicitly. Researchers should also test whether simpler semiparametric models can recover most inclusion benefits with lower implementation burden, because the best model is not always the most complex one available. That comparative discipline is essential for credible translation into regulated practice. and policy.

References

[1]. Ibrahim, M., Razib, M. N. H., Jahan, N., & Rahman, M. M. (2022). Climate risk, financial stability, and global capital allocation: A predictive analytics approach to assessing climate-related financial risk in international investment markets. Journal of Business and Management Studies, 4(4), 264–276. https://doi.org/10.32996/jbms.2022.4.4.34

[2]. Fahim, A. S. M., Ibrahim, M., Pritty, A. A., & Tania, T. A. (2023). Algorithmic accountability in U.S. consumer FinTech: Governance mechanisms for credit risk, fair lending, and financial stability. Journal of Economics, Finance and Accounting Studies, 5(4), 80–93. https://doi.org/10.32996/jefas.2023.5.4.8

[3]. Hasan, M. N., Rasel, I. H., Arman, M., Ibrahim, M., & Jahan, N. (2023). Strengthening U.S. financial and cybersecurity infrastructure with AI-driven fraud detection and risk analytics. Journal of Computational Analysis and Applications, 31(2), 15–32. Retrieved from eudoxuspress.com/index.php/pub/article/view/3823

[4]. Agarwal, S., Amromin, G., Ben-David, I., Chomsisengphet, S., & Evanoff, D. D. (2017). Policy intervention in debt renegotiation: Evidence from the Home Affordable Modification Program. Journal of Political Economy, 125(3), 654–712.

[5]. An, X., Cordell, L., Geng, L., & Lee, K. (2021). Default option exercise over the financial crisis and beyond. Review of Finance, 25(1), 153–187.

[6]. Avery, R. B., Brevoort, K. P., & Canner, G. B. (2007). Opportunities and issues in using alternative data in underwriting and credit scoring. Federal Reserve Bulletin, 93, A1–A20.

[7]. Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. Journal of the Operational Research Society, 54(6), 627–635.

[8]. Bellotti, T., & Crook, J. (2009). Support vector machines for credit scoring and discovery of significant features. Expert Systems with Applications, 36(2), 3302–3308.

[9]. Bellotti, T., & Crook, J. (2013). Forecasting and stress testing credit card default using dynamic models. International Journal of Forecasting, 29(4), 563–574.

[10]. Bequé, A., Coussement, K., Gayler, R., & Lessmann, S. (2017). Extreme learning machines for credit scoring: An empirical evaluation. Expert Systems with Applications, 86, 42–53.

[11]. Bhutta, N., Hizmo, A., & Ringo, D. (2021). How much does racial bias affect mortgage lending? Evidence from human and algorithmic credit decisions. Finance and Economics Discussion Series, Board of Governors of the Federal Reserve System.

[12]. Breeden, J. L. (2010). Reinventing retail lending analytics: Forecasting, stress testing, and scenario analysis. Risk Books.

[13]. Consumer Financial Protection Bureau. (2023). 2022 HMDA data on mortgage lending now available.

[14]. Consumer Financial Protection Bureau. (2023). Summary of 2022 data on mortgage lending.

[15]. Crook, J. N., Edelman, D. B., & Thomas, L. C. (2007). Recent developments in consumer credit risk assessment. European Journal of Operational Research, 183(3), 1447–1465.

[16]. Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. Applied Soft Computing, 91, 106263.

[17]. Demyanyk, Y., & Van Hemert, O. (2011). Understanding the subprime mortgage crisis. Review of Financial Studies, 24(6), 1848–1880.

[18]. Ding, L., Hwang, J., & Divringi, E. (2016). Gentrification and residential mobility in Philadelphia. Regional Science and Urban Economics, 61, 38–51.

[19]. Dumitrescu, E. I., Hué, S., Hurlin, C., & Tokpavi, S. (2022). Machine learning for credit scoring: Improving logistic regression with nonlinear decision-tree effects. European Journal of Operational Research, 297(3), 1178–1192.

[20]. Elul, R., Souleles, N. S., Chomsisengphet, S., Glennon, D., & Hunt, R. (2010). What triggers mortgage default? American Economic Review, 100(2), 490–494.

[21]. Ferreira, F., Gyourko, J., & Tracy, J. (2010). Housing busts and household mobility. Journal of Urban Economics, 68(1), 34–45.

[22]. Foote, C. L., Gerardi, K., & Willen, P. S. (2008). Negative equity and foreclosure: Theory and evidence. Journal of Urban Economics, 64(2), 234–245.

[23]. Foote, C. L., Gerardi, K., Goette, L., & Willen, P. S. (2010). Reducing foreclosures: No easy answers. NBER Macroeconomics Annual, 24(1), 89–138.

[24]. Fuster, A., & Willen, P. S. (2017). Payment size, negative equity, and mortgage default. American Economic Journal: Economic Policy, 9(4), 167–191.

[25]. Gerardi, K., Herkenhoff, K. F., Ohanian, L. E., & Willen, P. S. (2018). Can't pay or won't pay? Unemployment, negative equity, and strategic default. Review of Financial Studies, 31(3), 1098–1131.

[26]. Ghent, A. C., & Kudlyak, M. (2011). Recourse and residential mortgage default. Review of Financial Studies, 24(9), 3139–3186.

[27]. Gupta, A., Hansman, C., & Frenchman, D. (2022). Selection, leverage, and default in the mortgage market. Review of Financial Studies, 35(2), 720–767.

[28]. Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: A review. Journal of the Royal Statistical Society: Series A, 160(3), 523–541.

[29]. Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. Journal of Banking & Finance, 34(11), 2767–2787.

[30]. Li, W., Goodman, L., Bai, B., & Zhu, J. (2020). Measuring mortgage credit accessibility. Urban Institute.

[31]. Yıldırım, M., Uzgören, N., & Uzgören, E. (2021). Big data analytics for default prediction using graph theory. Expert Systems with Applications, 186, 115838.

[32]. Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research, 247(1), 124–136.

[33]. Liñares-Zegarra, J., & Wilson, J. O. S. (2018). The potential for fintech to improve credit access and financial inclusion. Journal of Financial Perspectives, 6(3), 1–15.

[34]. Longitudinal Employer-Household Dynamics Program. (2023). Data: LEHD Origin-Destination Employment Statistics (LODES). U.S. Census Bureau.

[35]. Mayer, C., Pence, K., & Sherlund, S. M. (2009). The rise in mortgage defaults. Journal of Economic Perspectives, 23(1), 27–50.

[36]. Munnell, A. H., Tootell, G. M. B., Browne, L. E., & McEneaney, J. (1996). Mortgage lending in Boston: Interpreting HMDA data. American Economic Review, 86(1), 25–53.

[37]. Paleologo, G., Elisseeff, A., & Antonini, G. (2010). Subagging for credit scoring models. European Journal of Operational Research, 201(2), 490–499.

[38]. Ross, S. L., & Yinger, J. (2002). The color of credit: Mortgage discrimination, research methodology, and fair-lending enforcement. MIT Press.

[39]. Scharlemann, T. C., & Shore, S. H. (2016). The effect of negative equity on mortgage default: Evidence from HAMP's principal reduction alternative. Review of Financial Studies, 29(10), 2850–2883.

[40]. Schulhofer-Wohl, S. (2011). Negative equity does not reduce homeowners' mobility. NBER Working Paper No. 16701.

[41]. Shumovskaia, V., Belyaev, A., Drutsa, A., & Katsev, P. (2021). Linking bank clients using graph neural networks powered by rich transactional data. Frontiers in Big Data, 4, 664758.

[42]. Silva, D. M. B., e Silva, A. A., & Marins, F. A. S. (2022). A class of categorization methods for credit scoring models. European Journal of Operational Research, 303(3), 1165–1179.

[43]. Thomas, L. C., Crook, J. N., & Edelman, D. B. (2002). Credit scoring and its applications. SIAM.

[44]. U.S. Census Bureau. (2023). OnTheMap and LEHD technical documentation.

[45]. Federal Housing Finance Agency. (2023). U.S. House Price Index report: 2022Q4.

[46]. Federal Housing Finance Agency. (2023). FHFA House Price Index methodology.

[47]. Xia, Y., Liu, C., Da, B., & Xie, F. (2017). A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. Expert Systems with Applications, 78, 225–241.

[48]. Yao, J., Crook, J., & Andreeva, G. (2015). Support vector machines for survival analysis in credit scoring. Operational Research, 15(2), 241–259.

[49]. Shumovskaia, V., Belyaev, A., Drutsa, A., & Katsev, P. (2021). Linking bank clients using graph neural networks powered by rich transactional data. Frontiers in Big Data, 4, 664758.