Construction and Building Materials 549 (2026) 148751

Contents lists available at ScienceDirect

Construction and Building Materials

journal homepage: www.elsevier.com/locate/conbuildmat

Effects of mixture design and cooling rate on the low-temperature TSRST responses of asphalt mixtures: Experimental evaluation and artificial neural network modelling

Adnan Qadir [a,1], Murat Guler [b,2], Yalcin Karakaya [c,*,3]

[a] Department of Urban and Infrastructure Engineering, NED University of Engineering and Technology, Karachi 75270, Pakistan
[b] Department of Civil Engineering, Middle East Technical University, Ankara 06800, Türkiye
[c] Department of Civil Engineering, Manisa Celal Bayar University, Manisa, 45140, Türkiye

ARTICLE INFO

Keywords:
Low-temperature cracking
Asphalt mixtures
Thermal Stress Restrained Specimen Test
Fracture strength
Fracture temperature
Thermal stress relaxation
Artificial neural network

ABSTRACT

This study evaluated seven material, mixture design, specimen, and testing factors within a Thermal Stress Restrained Specimen Test (TSRST) framework. A 72-combination fractional factorial design comprising 72 compacted slabs and two separately tested beam specimens from each slab produced 144 TSRST results for fracture strength (FS), fracture temperature (FT), relaxation stress (RS), and transition temperature (TT). Artificial neural network (ANN) models were fitted to the treatment combination means and evaluated using repeated nested cross-validation, grouped validation, and benchmark models. Aggregate type and binder content were the only factors significantly associated with all four TSRST responses. Limestone mixtures exhibited higher FS and RS and more negative FT and TT than basalt mixtures; however, because air void content differed substantially between aggregate groups, these results should be interpreted as differences between the investigated mixture systems rather than aggregate mineralogy alone. Mixtures prepared at their mixture-specific optimum binder content generally exhibited a more favourable balance of TSRST responses than mixtures deviating from the optimum. Styrene–butadiene–styrene (SBS) modification was associated with FS and FT but did not show consistent improvement across responses, whereas cooling rate was not significantly associated with any of the responses within the investigated −5 to −20 °C/h range. ANN performance was moderate for FS and RS, but limited for FT and TT, and benchmark comparisons showed that the ANN had no consistent predictive advantage over ridge or random forest regression. Overall, the study integrates fracture and pre-fracture TSRST assessment with preliminary ANN-based prediction within the experimental domain investigated.

1. Introduction

Low-temperature cracking is a major form of deterioration in asphalt pavements exposed to severe winter conditions and large temperature fluctuations. As the pavement temperature decreases, the bituminous layer undergoes thermal contraction. Because this contraction is restrained by the pavement structure and its interaction with the underlying layers, tensile thermal stresses develop in the asphalt mixture. When stress relaxation is insufficient and the accumulated thermal stress reaches the tensile strength of the mixture, cracking can initiate, typically in the transverse direction [1,2]. These cracks allow water to infiltrate, contribute to secondary pavement damage, reduce ride quality, and shorten pavement service life.

The Thermal Stress Restrained Specimen Test (TSRST), developed and refined through the Strategic Highway Research Program and subsequently issued as AASHTO TP 10–93, evaluates the development of thermally induced tensile stress in an asphalt mixture specimen that is restrained against axial contraction during controlled cooling. In the TSRST, the specimen is maintained at a constant length while its temperature is reduced at a specified rate. Thermal stress is monitored continuously until fracture, yielding fracture strength (FS) and fracture temperature (FT) as the main output parameters [3]. The resulting

* Corresponding author.
E-mail addresses: adnan@neduet.edu.pk (A. Qadir), gmurat@metu.edu.tr (M. Guler), yalcin.karakaya@cbu.edu.tr (Y. Karakaya).
[1] ORCID ID: 0000-0002-5088-9341
[2] ORCID ID: 0000-0001-6939-3322
[3] ORCID ID: 0000-0003-1036-7941

https://doi.org/10.1016/j.conbuildmat.2026.148751
Received 4 August 2026; Received in revised form 12 October 2026; Accepted 30 October 2026
Available online 2 November 2026
0950-0618/© 2026 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

stress–temperature curve generally comprises a relaxation-dominated region at higher temperatures and an approximately linear, non-relaxation-dominated region at lower temperatures. The transition temperature (TT) marks the boundary between these regions. Because contraction and relaxation properties influence thermal stress development during cooling, they provide important information for evaluating the low-temperature performance of asphalt mixtures [4]. In the present study, TT was operationally defined as the temperature at the beginning of the selected terminal linear segment while relaxation stress (RS) was defined as the fitted thermal stress at the same point. TT and RS were used to characterise the pre-fracture TSRST response.

Although the TSRST procedure is well established, the combined effects of material composition, mixture design, specimen geometry and cooling conditions on its response are still not fully characterised. Previous studies have typically examined individual variables or limited combinations, including aggregate characteristics [5], polymer modification [6], binder grade and content [7], and cooling conditions [8]. To the authors' knowledge, the combined effects of aggregate type, gradation, specimen length, binder grade, styrene–butadiene–styrene (SBS) modification, binder content, and cooling rate have rarely been examined within a single experimental study. Specimen length was included because previous experimental and numerical studies indicate that specimen geometry can influence thermal equilibration, stress development and TSRST fracture responses [9,10]. Binder content was evaluated at OAC − 0.5%, OAC, and OAC + 0.5%, as low-temperature mixture performance is governed by the combined effects of binder characteristics and binder content [7].

Machine learning methods, including artificial neural networks (ANN), are increasingly used to predict asphalt mixture properties and performance from material, mixture design, volumetric, and testing variables [11,12]. Previous studies have demonstrated their potential for predicting performance indicators such as Hamburg wheel-tracking rut depth and indirect tensile strength [11], as well as low-temperature fracture resistance curves [12]. By capturing complex, potentially nonlinear relationships among multiple input variables, these methods can support preliminary evaluation of mixtures and screening of performance. However, their predictive reliability depends on the quantity, quality and representativeness of the training data, as well as on appropriate model validation [11,12]. Machine learning predictions should therefore be regarded as complementary to, rather than replacements for, laboratory performance testing.

The present study addresses these research gaps through analysis of a structured multifactor experimental programme. The principal contribution is not the development of a new TSRST procedure, but the simultaneous evaluation of material, mixture design, specimen, and testing variables within a common experimental framework. A total of 72 treatment combinations, each represented by two separately tested within-slab beam specimens, were analysed, giving 144 TSRST specimens in total. Seven experimental factors were considered: aggregate type, gradation, specimen length, SBS modification, binder penetration grade, binder content and cooling rate. Unlike comparisons based on separate single-factor studies using different materials and test conditions, this design provides a common basis for assessing both the relative and response-specific influence of these factors on fracture responses (FS and FT) and pre-fracture responses (RS and TT). The objectives were to: (i) quantify the main effects of these factors on fracture strength (FS), fracture temperature (FT), relaxation stress (RS), and transition temperature (TT); (ii) determine whether the factors influencing fracture responses also influence pre-fracture stress development and relaxation; and (iii) apply separate artificial neural network models to estimate TSRST responses from mixture design variables, test conditions, and mean air void content, evaluate their performance using internal cross-validation, and compare their predictive accuracy with ridge regression and random forest benchmark models.

The experimental programme analysed in this study originates from the broader TSRST campaign documented in the doctoral work of Qadir [13]. Karakaya et al. [14] previously analysed a selected subset of 48 of these 144 TSRST observations comprising 300-mm specimens tested at −10 °C/h. That study used the 48-observation TSRST subset to evaluate fracture strength and fracture temperature, together with separate thermal fatigue and direct tension datasets. The present study extends the analysis to the full 144-observation TSRST dataset, comprising 72 treatment combinations across the investigated specimen lengths and cooling rates, and additionally evaluates pre-fracture relaxation stresses and transition temperature responses. Thus, 48 TSRST observations are shared between the two studies, while the remaining 96 TSRST observations, together with the broader seven-factor TSRST analysis, are included only in the present study.

Accordingly, the contribution of the present study lies primarily in the integrated analysis of the broader, pre-existing TSRST dataset, the response-level comparison of fracture and pre-fracture behaviour, and the cross-validated application of ANN models, rather than in the generation of new experimental data or the introduction of new testing, data-processing, or machine learning methodologies. The present analysis examines how the seven investigated factors are associated with FS, FT, RS, and TT across the broader TSRST test matrix and evaluates the extent to which these responses can be estimated using ANN models assessed through repeated nested cross-validation and a grouped cross-validation sensitivity analysis. The fractional factorial design enables the associations of the seven factors with TSRST responses to be examined under consistent specimen preparation and testing conditions while the simultaneous analysis of FS, FT, RS, and TT distinguishes factor-associated differences in fracture responses from those observed during pre-fracture thermal stress development. Table 1 summarises the relationships between the datasets, previously reported results, analyses, figures and scientific scope of the two studies.

Table 1
Relationship between the previously published study [14] and the present study.

Aspect	Karakaya et al. [14]	Present study
TSRST dataset	Selected 48-specimen subset of the broader programme; 300-mm specimens tested at −10 °C/h	Complete 144-specimen TSRST dataset representing 72 treatment combinations across 200/300 mm and −5/−10/−20 °C/h
Experimental factors	Five mixture-design factors: aggregate type, gradation, SBS modification, binder grade, and binder content	Seven factors: the same five mixture-design factors plus specimen length and cooling rate
Outcomes and data overlap	FS and FT reported for the selected 48 TSRST specimens; separate TF and DTT datasets also analysed	Includes the previously reported FS and FT results for those 48 specimens plus FS/FT for 96 additional TSRST specimens; RS and TT evaluated for all 144 specimens
Statistical analyses	Five-factor analysis of the selected LTC dataset, together with separate TF and DTT analyses	Seven-factor analysis of the complete TSRST dataset and air-void-response analyses
Predictive modelling	None	Cross-validated ANN modelling with ridge and random-forest benchmark models
Figures	LTC factor-response figures and separate TF/DTT figures	All figures are newly prepared for the present study; no figure from [14] is reproduced.
Scientific scope	Comparison of factor sensitivities across LTC, TF, and DTT response modes	Integrated evaluation of fracture and pre-fracture TSRST responses across seven factors, together with predictive modelling

2. Literature review

2.1. Low-temperature cracking mechanisms in asphalt pavements

Low-temperature cracking in asphalt pavements is governed by the balance between thermal stress development and viscoelastic stress relaxation. At temperatures above the glass transition range of the asphalt binder, thermally induced stresses can relax more readily. As the temperature approaches or falls below this range, stress relaxation slows and the material stiffens, reducing the mixture's capacity to dissipate thermal stress [15,16]. Near the glass-transition range, thermal stress development remains strongly time- and temperature-dependent and is influenced by the cooling rate and prior low-temperature conditioning [15]. Consequently, tensile thermal stress accumulates, and fracture may occur when it reaches the mixture's tensile strength [17].

Christison et al. [18] presented an early mechanistic evaluation of low-temperature cracking by estimating thermally induced stresses in asphalt concrete pavement sections, establishing fracture strength–temperature relationships from pavement cores, and comparing predicted fracture susceptibility with observed field cracking. Subsequent advances in asphalt rheology provided broader methods for characterising the temperature- and time-dependent responses of binders and mixtures. Zeng et al. [19], for example, developed a unified rheological model that describes modified asphalt binders as viscoelastic fluids and asphalt mixtures as viscoelastic solids over relevant ranges of temperature, loading frequency, and strain. Collectively, these studies demonstrate the importance of considering both thermally induced stress development and the temperature-dependent mechanical response of the asphalt mixture when evaluating low-temperature cracking.

At the mixture level, low-temperature cracking resistance cannot be inferred from binder rheology alone, because aggregate characteristics and volumetric properties also influence the fracture response of the mixture. Jung and Vinson [9] found that TSRST fracture temperature was most strongly influenced by degree of ageing and asphalt type, whereas fracture strength was most strongly influenced by air void content and aggregate type. Using semi-circular bending tests, Li et al. [20] further demonstrated that low-temperature fracture resistance was strongly dependent on test temperature and significantly influenced by aggregate type, air void content, binder grade, and binder modifier type. Binder content significantly affected fracture energy but did not significantly affect fracture toughness [20]. These findings suggest that fracture strength and fracture temperature represent related but distinct aspects of the TSRST response and may display different sensitivities to mixture design variables.

2.2. The TSRST: Principles and prior applications

Jung and Vinson [9] reported on the development and systematic evaluation of the Thermal Stress Restrained Specimen Test (TSRST) for assessing the resistance of asphalt mixtures to low-temperature cracking. Unlike binder-level rheological tests, the TSRST characterises mixture-level behaviour and its response reflects the combined influence of binder properties, aggregate structure, volumetric characteristics, thermal contraction, stress relaxation, and fracture resistance of the mixture. During controlled cooling, the specimen is restrained from contracting and the resulting thermal stress is monitored continuously until fracture. The test therefore provides fracture strength and fracture temperature as its principal outputs [9] while the pre-fracture stress–temperature curve offers additional information on thermal stress development and relaxation [4,9]. In the present study, transition temperature and relaxation stress were retained as supplementary curve-derived indicators.

Falchetto et al. [21] investigated whether low-temperature properties obtained from Bending Beam Rheometer creep and strength tests on small asphalt mixture specimens could be related to those measured using TSRST. For eight mixtures incorporating different combinations of conventional and recycled materials, the critical cracking temperatures initially obtained from the two methods differed significantly. A simplified size-effect correction was therefore applied to convert the bending beam rheometer (BBR)-measured strength to an equivalent tensile strength corresponding to the larger TSRST specimens. Although this correction improved the correspondence between the two methods, the resulting critical cracking temperatures were still only partly consistent. The study therefore showed that direct comparison between small-specimen BBR and TSRST results requires consideration of specimen size effects and that the results of the two methods should not be regarded as directly interchangeable.

The effects of binder characteristics, SBS modification, and cooling rate on TSRST response have also been investigated. Pszczola et al. [22] compared mixtures containing neat, SBS-modified, and highly SBS-modified binders and separately evaluated the influence of cooling rates from 1 °C/h to 10 °C/h on a mixture containing neat 35/50 binder. They found that binder type had the greatest influence on the TSRST results, followed by cooling rate and the additives investigated. Increasing the cooling rate generally shifted fracture towards higher temperatures, indicating a less favourable low-temperature response [22]. Using analytical viscoelastic calculations, finite-element simulations and pavement-cracking predictions, Apeagyei et al. [23] similarly demonstrated that faster cooling increases thermal stress accumulation because less time is available for viscoelastic relaxation. However, different mixtures exhibited varying sensitivities to cooling rate, and the magnitude of this effect became less pronounced at higher cooling rates [23]. These findings indicate that cooling rate is an important testing and environmental variable, but its influence should be interpreted alongside binder characteristics and mixture composition.

The nominal cooling rate applied to the chamber air may not accurately represent the actual cooling rate at the specimen core. To reduce this source of discrepancy in the current experimental programme, the cooling control system used temperature feedback from four resistance temperature detectors (RTDs) mounted on the specimen surfaces rather than relying solely on the chamber air temperature. Jung and Vinson [9] identified a significant specimen-size effect on TSRST fracture temperature when comparing specimens with different overall dimensions; however, specimen length and cross-sectional dimensions varied simultaneously, preventing the independent effect of length from being isolated. Specimen size was not statistically significant for fracture strength, although differences in the mean values were observed. Steiner and Hofko [24] demonstrated that climatic chamber configuration, airflow velocity, and specimen cross-section can influence the temperature lag between the chamber air and the specimen core. Greater thermal lag slows specimen cooling and thermal stress accumulation, allows more time for viscoelastic relaxation, and can produce systematic differences in measured fracture temperatures between TSRST devices. Consequently, specimen core temperature, chamber airflow, and specimen dimensions should be considered when interpreting and comparing TSRST results.

Despite these developments, relatively few studies have jointly considered material, mixture design, specimen, and testing factors within a structured experimental programme. In particular, the combined effects of aggregate type, gradation, specimen length, binder penetration grade, SBS modification, binder content, and cooling rate on both fracture and pre-fracture TSRST responses remain insufficiently characterised. The present study addresses this gap by evaluating these factors within a common experimental framework and by considering fracture strength and fracture temperature, together with transition temperature and relaxation stress. This broader assessment enables the relative influence of the investigated variables on different stages of the TSRST response to be examined and provides an experimental basis for the subsequent statistical and machine learning analyses.

2.3. Influence of mixture design variables on low-temperature performance

Aggregate type can significantly influence the low-temperature fracture response of asphalt mixtures. However, its effect should be interpreted in conjunction with binder properties, mixture composition, volumetric characteristics, test temperature and the fracture parameter under consideration. Li et al. [20] found that aggregate type and air void content significantly affected the low-temperature fracture resistance of mixtures evaluated using semi-circular bending tests; binder grade and binder modification were also significant factors. Braham et al. [25] similarly reported that aggregate type and test temperature significantly influenced the fracture energy measured in disk-shaped compact tension tests. Using asymmetric semi-circular bend specimens, Aliha et al. [26] further investigated the effects of aggregate type and size, binder type, and air void content on mixed-mode low-temperature fracture toughness and showed that these mixture characteristics influenced the measured fracture response. These findings support aggregate type as an important mixture-level factor, although its influence depends on the other material, volumetric, and testing conditions considered.

Aggregate gradation influences the internal structure of asphalt mixtures and can alter stress transfer and the paths available for crack propagation. Therefore, its effect on low-temperature performance cannot be expressed as a consistent advantage of one gradation type under all mixture and testing conditions. Using virtual indirect tensile tests, Du et al. [27] compared numerical models of dense-, open- and gap-graded asphalt mixtures and showed that gradation affected mixed-mode crack propagation, crack-tip stress-field distributions and fracture energy release behaviour under low-temperature loading. Xiao et al. [28] further demonstrated that railway asphalt concrete mixes with different aggregate gradations exhibited distinct crack trajectories and fracture responses in low-temperature semi-circular bending tests. These findings indicate that aggregate gradation alters the mixture structure and low-temperature fracture behaviour [27,28]. Its influence should therefore be interpreted alongside other relevant factors, including binder characteristics, binder content and air void content [20,28].

Styrene–butadiene–styrene (SBS) modification is widely used to alter the rheological properties of asphalt binder. Its effectiveness depends on formulation-related factors, including the properties and composition of the base binder, SBS type and content, polymer dispersion and compatibility, and the presence of additional modifiers or processing aids [6,29]. Dong et al. [29] showed that base asphalt composition, SBS structure, and SBS content significantly influence polymer dispersion and the resulting properties of SBS-modified binders. Lin et al. [6] further demonstrated that SBS type, SBS content, sulphur content, and rubber-processing oil influence low-temperature rheological properties and relaxation capacity. Their TSRST results showed that mixture-level low-temperature performance was sensitive to the chosen modification formulation and supported the identification of an optimum formulation. Pszczola et al. [22] similarly observed different TSRST responses among mixtures containing neat, SBS-modified, and highly SBS-modified binders and separately demonstrated that cooling rate influences TSRST results. Collectively, the available findings indicate that the low-temperature response associated with SBS modification may vary with the binder formulation, mixture composition, and testing conditions [6,22,29].

Binder content is closely related to effective binder content, asphalt film thickness, and mixture volumetrics, which are associated with low-temperature fracture performance [30]; it can also influence thermal stress development [31]. However, its observed influence is not uniform and may vary with the response parameter, test temperature, aggregate type, and volumetric conditions [20,25,30]. Rys et al. [7] reported that the TSRST failure temperature depended on binder rheological and brittle-fracture characteristics, together with binder content while Akentuna et al. [31] found that binder content and air void content affected thermal stress development measured using the Asphalt Concrete Cracking Device (ACCD). Fracture studies have similarly shown that binder content can significantly affect fracture energy without producing a correspondingly significant effect on fracture toughness and that its influence on fracture energy may vary with test temperature [20,25]. Binder content should therefore be interpreted in conjunction with binder characteristics and mixture volumetrics rather than as an isolated design variable.

Overall, previous research indicates that aggregate type and gradation, binder penetration grade and SBS modification, binder content, and volumetric properties collectively influence the low-temperature behaviour of asphalt mixtures. Their effects may vary in terms of fracture strength, fracture temperature, thermal stress relaxation, and crack-propagation parameters, indicating that no single material or mixture design variable is sufficient to fully characterise low-temperature performance. However, the studies reviewed above have generally examined individual factors or limited combinations under specific test conditions. This limitation underscores the need for a structured multivariable evaluation that considers material, mixture design, specimen, and testing factors within a common experimental framework. The present study addresses this need by examining their effects on both fracture and pre-fracture TSRST responses.

2.4. Artificial intelligence and machine learning in asphalt mixture design

Machine learning methods have been applied to predict asphalt mixture properties and performance from material, mixture design, volumetric, and testing variables [11,32]. Rahman et al. [11] used support vector regression and tree-based ensemble methods to predict Hamburg rut depth and indirect tensile strength from aggregate, binder, recycled material, density, and test-related variables. Their findings showed that machine learning models can support performance-oriented mixture design and evaluation when trained on sufficient, carefully curated data. More broadly, the systematic review by Leukel et al. [32] documented the adoption of machine learning methods in asphalt road construction while identifying limitations related to the comparability of evaluation metrics, the reporting of training and test data, and the assessment of model generalisability.

Machine learning methods have also been applied specifically to low-temperature fracture performance. Mirzaiyanrajeh et al. [33] developed artificial neural network and self-validated ensemble models to predict the low-temperature fracture energy of asphalt mixtures using a database of 852 fracture energy values obtained from disk-shaped compact tension (DCT) tests. Their findings showed that variables available during mixture design can be used to obtain preliminary estimates of fracture energy, and a self-validated ensemble model was proposed as a predesign tool for situations in which laboratory testing is not feasible. Ghafari et al. [12] used a multilayer perceptron artificial neural network and multi-gene genetic programming to predict low-temperature fracture resistance curves using aggregate type, binder content, nominal maximum aggregate size, binder performance grade, crumb rubber content, and test temperature as input variables. Although these studies used fracture tests other than the TSRST, they demonstrate the potential of machine learning models to represent complex relationships between mixture characteristics and low-temperature fracture responses.

A practical challenge in applying machine learning methods to TSRST responses is obtaining sufficiently large and diverse experimental datasets because TSRST specimen preparation and testing are time- and resource-intensive. Model generalisability is influenced by the size and diversity of the available dataset while reported predictive performance can be affected by the validation strategy and evaluation procedures used [32,34]. In the present study, 144 specimens represented 72 distinct treatment combinations across seven experimental factors. Although the replicate specimens provided information on within-combination variability, coverage of the experimental factor space was determined primarily by the 72 distinct treatment

combinations. The ANN models were therefore fitted at the treatment–combination level and internally evaluated using repeated nested cross-validation, with model selection performed independently within the training set. A grouped validation was also used to evaluate performance when complete underlying mixture formulations were excluded from the training set. These procedures provide an internal assessment of performance for unseen combinations within the materials and experimental domain investigated, but they do not allow extrapolation to new materials or experimental factor settings. Previous machine learning studies have demonstrated the potential of data-driven approaches for estimating the low-temperature fracture responses of asphalt mixtures [12,33]. In the present study, the ANN component is used as a complementary analysis of the multifactor TSRST dataset rather than as a new machine learning method. Its specific contribution is the internal cross-validation of ANN-based estimates for four TSRST responses, using the investigated mixture, volumetric, specimen, and testing variables, including the pre-fracture indicators RS and TT.

3. Materials and methods

3.1. Experimental design

A fractional factorial experimental design was used to evaluate the effects of seven experimental factors on the TSRST responses. Five factors were investigated at two levels: aggregate type, gradation, specimen length, SBS modification, and binder penetration grade. Binder content and cooling rate were investigated at three levels. The corresponding full factorial design comprised $2^5 \times 3^2 = 288$ possible treatment combinations. To reduce the experimental demand while retaining structured coverage of the investigated factor space, 72 unique treatment combinations were selected, representing a one-quarter fraction of the full design. The realised 72-combination matrix was non-regular and therefore did not have the simple, complete aliasing chains found in regular two-level fractional factorial designs. Partial non-orthogonality nevertheless remained, so contributions from unmodelled interactions to the estimated main effects cannot be ruled out. One compacted slab represented each of the 72 treatment combinations. Two prismatic beam specimens were taken from each slab and tested separately, giving a total of 144 TSRST specimens. Thus, the two specimens associated with each treatment combination were replicate test specimens taken from within the same slab, rather than independently prepared batch or slab replicates. The cooling rate was balanced across the selected combinations, with 24 treatment combinations (48 specimens) at each cooling rate. Because the 72-combination matrix represented a selected one-quarter fraction of the 288-combination full factorial design, the marginal numbers of combinations were not equal across all levels of the remaining factors. This imbalance therefore reflects the structure of the selected fractional design rather than missing or excluded specimens; two valid within-slab specimen results were available for each of the 72 treatment combinations. The resulting level-specific sample sizes are reported in the Results. The factors investigated, their levels, and the notation used throughout the study are summarised in Table 2.

The 144 TSRST observations analysed here form part of the experimental programme originally documented by Qadir [13]. Karakaya et al. [14] previously analysed a selected subset comprising 48 of the 144 TSRST observations, specifically the 300-mm specimens tested at −10 °C/h. The present analysis uses the broader TSRST matrix and pursues a different objective by jointly evaluating seven experimental factors, four TSRST response variables, and their data-driven estimation using ANN models.

Table 2
Experimental factors and levels.

Experimental factor	Number of levels	Investigated levels
Aggregate type	2	Limestone / Basalt
Gradation	2	Coarse / Fine
Specimen length	2	200 / 300 mm
SBS modification	2	Unmodified / SBS-modified
Binder grade	2	50/70 and 70/100 penetration grade
Binder content	3	OAC − 0.5% / OAC / OAC + 0.5%
Cooling rate	3	−5 / −10 / −20 °C/h

Note: OAC = the optimum asphalt content. OAC − 0.5% and OAC + 0.5% denote deviations of − 0.5 and + 0.5 % from the mixture-specific OAC, respectively.

3.2. Materials

Limestone and basalt aggregates obtained from an asphalt plant near Ankara, Turkey, were used in the experimental programme. Two aggregate gradations were selected in accordance with the Turkish General Directorate of Highways specification for asphalt wearing courses. Type I represented the coarse gradation while Type II represented the fine gradation. The selected percentage-passing values are presented in Table 3. The limestone and basalt aggregates had specific gravities of 2.752 and 2.931, water absorption values of 1.45% and 1.00%, and Los Angeles abrasion values of 28% and 15%, respectively.

Two penetration-grade binders, 50/70 and 70/100, were investigated in unmodified and SBS-modified forms. The measured penetration values of the two base binders were 54 and 73 in units of 0.1 mm while their specific gravities were 1.025 and 1.034, respectively. The SBS-modified binders were supplied ready for use by the Turkish General Directorate of Highways; therefore, no laboratory modification was carried out by the authors, and the SBS dosage and production protocol were not available.

3.3. Mixture design

The optimum asphalt content (OAC) was determined separately for 16 base mixture designs representing the full factorial combination of two aggregate types, two gradations, two binder penetration grades, and two SBS-modification conditions $(2^4 = 16)$. The Superpave volumetric mix design procedure was applied using a design gyration level of 100, a design traffic level of 3–10 million equivalent single-axle loads (ESALs), and a target air void content of 4%. The mixing and compaction temperatures were 165 °C and 145 °C, respectively. Three specimens were initially prepared for each base mixture to determine the OAC, and additional specimens were subsequently prepared to confirm the selected value. A total of 96 cylindrical specimens were produced during the mixture design stage. The resulting OAC values ranged from 4.4% to 5.5%, as presented in Table 4. These mixture-specific OAC values were subsequently used to define the three binder content levels investigated in the TSRST programme: OAC − 0.5, OAC, and OAC + 0.5 %e.

3.4. Specimen preparation

One compacted slab was prepared for each treatment combination. Approximately 22–23 kg of asphalt mixture was required for each slab.

Table 3
Aggregate gradations used in the experimental programme.

Sieve	Sieve size (mm)	Type I, coarse gradation (% passing)	Type II, fine gradation (% passing)
3/4 in.	19.10	100	100
1/2 in.	12.50	83	100
3/8 in.	9.52	70	90
No. 4	4.76	40	72
No. 10	2.00	25	53
No. 40	0.420	10	28
No. 80	0.177	6	16
No. 200	0.075	4	10

Table 4
Optimum asphalt contents of the base mixtures.

Aggregate	Binder grade	Modification	Gradation	OAC (%)
Basalt	50/70	Unmodified	Fine	5.0
Basalt	50/70	Unmodified	Coarse	4.4
Basalt	50/70	SBS-modified	Fine	5.2
Basalt	50/70	SBS-modified	Coarse	4.8
Basalt	70/100	Unmodified	Fine	5.3
Basalt	70/100	Unmodified	Coarse	4.5
Basalt	70/100	SBS-modified	Fine	5.4
Basalt	70/100	SBS-modified	Coarse	5.0
Limestone	50/70	Unmodified	Coarse	4.5
Limestone	50/70	Unmodified	Fine	5.3
Limestone	50/70	SBS-modified	Coarse	5.0
Limestone	50/70	SBS-modified	Fine	5.5
Limestone	70/100	Unmodified	Coarse	4.8
Limestone	70/100	Unmodified	Fine	5.4
Limestone	70/100	SBS-modified	Coarse	5.1
Limestone	70/100	SBS-modified	Fine	5.5

Note: OAC = optimum asphalt content.

Because of the limited mixer capacity, the material required for a single slab was prepared in four sequential mixer batches, which together formed the mixture used for that slab; these batches were therefore not independent experimental replicates. Before compaction, the mixtures underwent three hours of oven conditioning to simulate short-term ageing. Then, the mixtures were compacted into slabs measuring 500 mm × 180 mm × 100 mm (length × width × thickness) using a French LCPC slab compactor, as shown in Fig. 1(a). The resulting compacted slabs are presented in Fig. 1(b). Following compaction, the slabs were cut using a water-cooled diamond saw, as shown in Fig. 1(c). Prismatic specimens with a constant cross-section of 50 mm × 65 mm and lengths of either 200 mm or 300 mm were obtained, as shown in Fig. 1(d). The 200 mm and 300 mm specimens were designated as short and long, respectively. Specimen length was included as an experimental factor to evaluate its association with the TSRST responses.

Each specimen was assigned a code indicating its mixture configuration and cutting position within the original slab. After cutting and dimensional finishing, the bulk specific gravity of each specimen was determined using the saturated-surface-dry procedure specified in AASHTO T 166. Air void content was calculated from the measured bulk specific gravity and the theoretical maximum specific gravity of the corresponding mixture, determined in accordance with AASHTO T 209. Two prismatic TSRST specimens were cut from each compacted slab, giving a total of 144 specimens from 72 slabs. The two specimens associated with each treatment combination therefore shared the same slab origin and should be regarded as within-slab replicates rather than independently prepared experimental units.

3.5. TSRST testing procedure

The TSRST programme was conducted using a purpose-built, computer-controlled servo system consisting of a mechanical loading assembly and a temperature-controlled environmental chamber. Prismatic specimens were bonded to steel loading platens with a high-strength epoxy adhesive and mounted vertically in the chamber. Axial displacement was monitored using two linear variable displacement transducers (LVDTs) positioned on opposite sides of the specimen while the specimen surface temperature was measured using four PT-100 Class A resistance temperature detectors. The prescribed cooling programme was controlled by the liquid nitrogen system, using temperature feedback from four RTDs attached to the specimen surfaces. The control software continuously displayed the prescribed and actual cooling rates, along with the measured specimen surface and chamber temperatures. Air circulation within the chamber was driven by fans, and liquid nitrogen was delivered through multiple nozzles to ensure uniform

Fig. 1. Preparation of prismatic TSRST specimens: (a) compaction using the French LCPC slab compactor; (b) compacted asphalt slabs; (c) cutting of the slabs using a diamond saw; and (d) final prismatic TSRST specimens.

cooling. The cooling system was further monitored using a dummy asphalt specimen fitted with an internal temperature sensor. These measurements were used to adjust the chamber control settings, achieve the prescribed specimen cooling rates, and reduce discrepancies between the imposed chamber conditions and the internal thermal response of the specimens. The arithmetic mean of the four RTD readings was used as the specimen temperature for constructing the stress–temperature response and determining the fracture temperature. The mounted TSRST specimen configuration, including the two LVDTs, surface-mounted RTDs and the specimen after fracture, is shown in Fig. 2. The servo-controlled loading assembly and the liquid-nitrogen supply system are illustrated in Fig. 3(a) and Fig. 3(b), respectively. Each specimen was conditioned at 5 °C for 3 h before testing. It was then cooled at the prescribed rate of −5, −10, or −20 °C/h. During cooling, the actuator compensated for thermally induced shortening and maintained the specimen's initial axial length, causing tensile stress to develop under restrained contraction. Axial load, specimen temperature and displacement were recorded at 5 s intervals until fracture.

The measured load was converted into thermal tensile stress using the measured specimen cross-sectional area. The stress–temperature records were processed using the MATLAB-based procedure associated with the original experimental programme [13]. This procedure is summarised below to support reproducibility. The stress and temperature records were smoothed separately using a 25-point moving average. Linear, quadratic, and cubic polynomial functions were then evaluated, and the polynomial associated with the highest calculated coefficient of determination (R^2) was selected to represent the overall stress–temperature response. Fracture temperature (FT) was defined as the minimum specimen temperature reached at failure and fracture strength (FS) was calculated from the selected polynomial fit at FT. Relaxation stress (RS) and transition temperature (TT) were determined using a separate linear-segment procedure. Linear regressions were fitted to four candidate terminal portions of the stress–temperature record, comprising the final 25%, approximately one-third, 50%, and 75% of the observations. The segment with the highest coefficient of determination was selected as the approximately linear part of the response. TT was defined as the temperature at the beginning of this segment, and RS was the fitted thermal stress at the same point, as shown in Fig. 4.

RS and TT were identified automatically using the same predefined procedure for all specimens, without manual selection of the transition point. Nevertheless, because these parameters depend on identifying the approximately linear terminal portion of the stress–temperature response, they are operationally defined quantities associated with the adopted processing procedure rather than direct, processing-independent material properties.

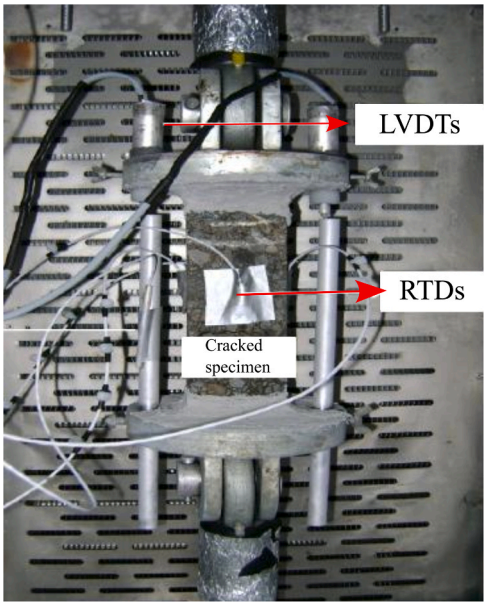

Fig. 2. TSRST specimen mounted inside the environmental chamber, showing the liquid-nitrogen spray nozzle, surface-temperature RTD, loading fixtures, and a representative fracture crack.

3.6. Statistical analysis

Statistical analyses were conducted at the specimen level using 144 valid TSRST test results. The dataset comprised 72 treatment combinations, each represented by two beam specimens cut from the same compacted slab. The two beams were physically distinct specimens and were tested separately, but they were not independent mixture-production or slab-level experimental replicates. Their individual responses were retained in the specimen-level ANOVA to preserve the observed within-slab variability. Accordingly, the 144 observations represent separate TSRST measurements from 72 slab-level treatment combinations. Because the statistical model does not explicitly account for the shared slab origin of the paired specimens, the reported significance levels should be interpreted with appropriate caution and in the context of this experimental design. Separate additive factorial ANOVA models were fitted for fracture strength (FS), fracture temperature (FT), relaxation stress (RS), and transition temperature (TT). Aggregate type, gradation, specimen length, SBS modification, binder penetration grade, binder content level, and cooling rate were entered as fixed categorical factors. Each factor was evaluated after adjustment for the other experimental factors included in the model. Statistical significance was assessed at $\alpha = 0.05$.

Because the realised fractional design was partially non-orthogonal, unmodelled interaction effects may have influenced the estimated main effects, even though no main effect was completely aliased with any two-factor interactions. The reported statistical significances and main-effect estimates should therefore be interpreted as factor–response associations conditional on the specific experimental design under investigation rather than as definitive standalone causal effects. Accordingly, this analysis is intended to identify statistically supported associations within the study design and does not establish mechanistic causal relations.

Pearson product–moment correlation coefficients and least-squares linear regression lines were used descriptively to examine the relationships between air void content and the TSRST responses at the specimen level. These relationships were examined for the complete dataset and, where relevant, within the aggregate-type and aggregate–modification subgroups shown in Figs. 5 and 6. In addition, a separate two-factor ANOVA was fitted to air void content with aggregate type, SBS modification, and their interaction included as fixed explanatory factors.

The correlation analyses were used only to describe the relationship between volumetric condition and TSRST response and were not intended to statistically isolate any intrinsic aggregate-related or SBS-related effects. Air void content was treated as a characteristic of the resulting mixture system rather than as an independently controlled covariate, because it may itself reflect aggregate characteristics, binder modifications, mixture composition, and compactability. Accordingly, no attempt was made to separate volumetric and material contributions causally.

3.7. Artificial neural network modelling, validation and benchmark comparison

Four separate feed-forward multilayer perceptron artificial neural network models were fitted to estimate fracture strength (FS), fracture temperature (FT), relaxation stress (RS), and transition temperature (TT). The experimental dataset comprised 144 specimen-level observations obtained from 72 compacted slabs/treatment combinations. The two within-slab specimen results for each treatment combination were

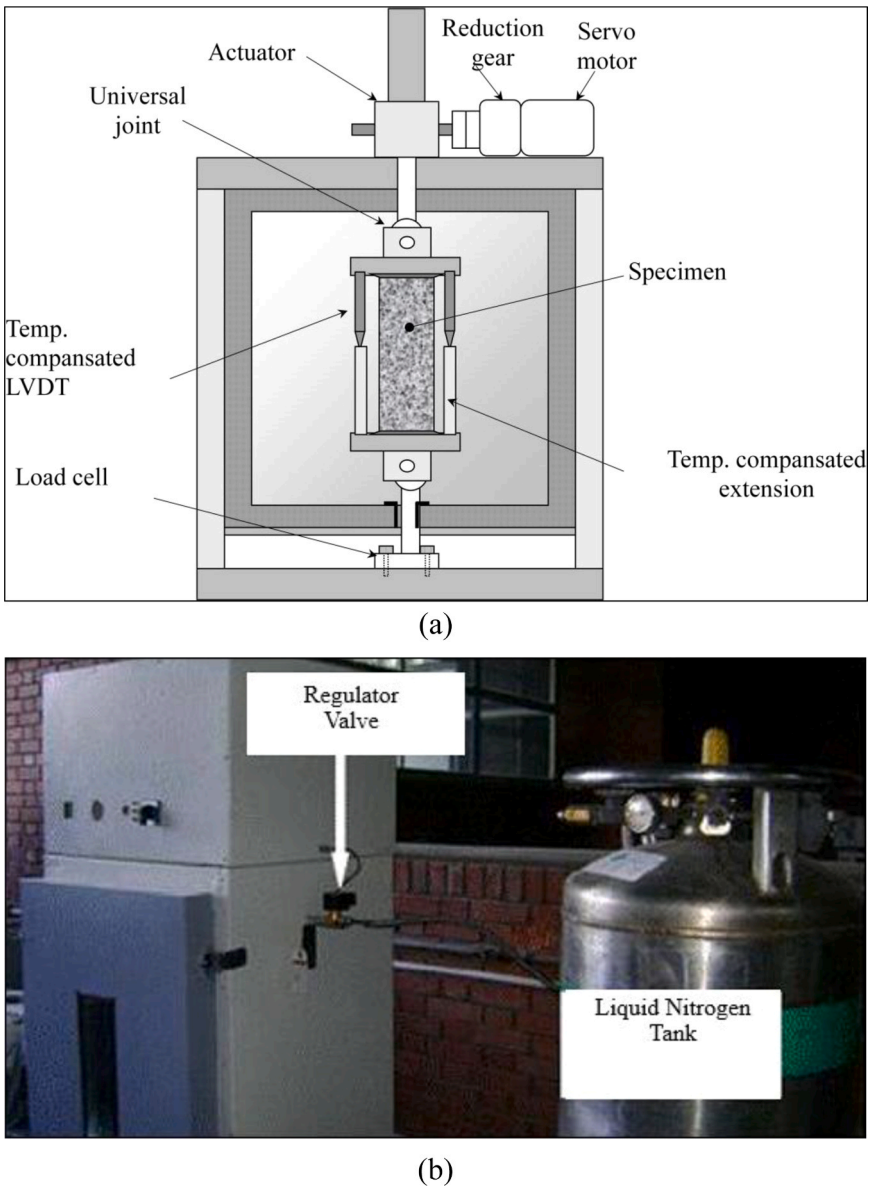

Fig. 3. Main components of the TSRST loading and cooling systems: (a) schematic of the servo-motor-driven loading assembly, including the reduction gear, actuator, and universal joint; and (b) liquid-nitrogen regulator valve and supply arrangement.

averaged before ANN modelling, resulting in 72 records.

Eight predictors were considered: aggregate type, gradation, specimen length, SBS modification, binder penetration grade, binder content level, cooling rate, and mean air void content. Aggregate type, gradation, specimen length, SBS modification, binder penetration grade, and binder content level were treated as categorical predictors and encoded using one-hot encoding. Cooling rate and mean air void content were treated as numerical predictors and standardised. The response variable was also standardised during model fitting. All encoding and scaling parameters were estimated exclusively from the relevant training data within each cross-validation fold.

Because only 72 distinct treatment combinations were available, a deliberately restricted common hyperparameter space was used for all four responses to limit complexity and reduce overfitting during model selection. Candidate hidden-layer configurations comprised one hidden layer with 4 neurons, one hidden layer with 8 neurons, and two hidden layers with 8 and 4 neurons, respectively. L2 regularisation parameters of 0.01, 0.1, and 1.0 were evaluated. All candidate networks used the hyperbolic tangent (tanh) activation function, the Adam optimiser, an initial learning rate of 0.001, automatic batch size selection, a maximum of 2000 iterations, a convergence tolerance of 10^{-3}, 20 iterations without sufficient improvement as the stopping criterion, and a random seed of 42.

Predictive performance was evaluated using repeated nested cross-validation. The outer validation consisted of fivefold cross-validation repeated 10 times. Within each outer training dataset, hyperparameter selection was performed independently using fourfold inner cross-validation, with root mean square error (RMSE) as the selection criterion. Consequently, observations in each outer test fold were not used for predictor preprocessing, response scaling, hyperparameter tuning, or model fitting. R^2, mean absolute error (MAE), and RMSE were calculated from the complete out-of-fold predictions for each repetition. The reported performance values represent the mean $\pm$ standard deviation across the 10 outer cross-validation repetitions.

To determine whether the ANN provided predictive benefit relative to alternative benchmark models, its performance was compared with ridge regression and random forest regression using the same outer folds in each of the 10 repeated fivefold cross-validation runs. Ridge regres-

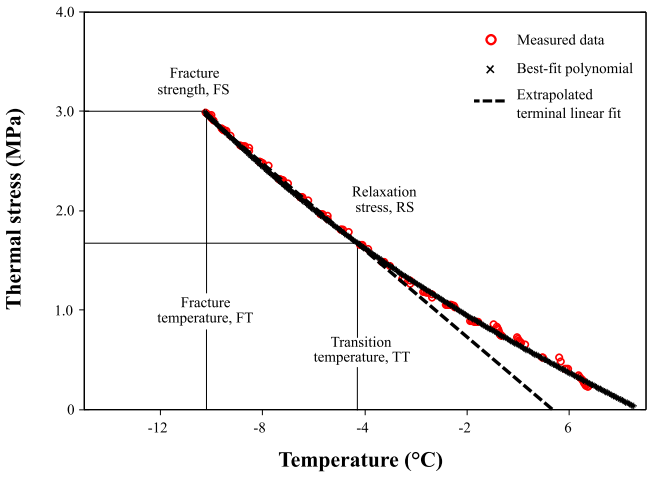

Fig. 4. Representative TSRST thermal stress–temperature response illustrating the extraction of fracture strength (FS), fracture temperature (FT), relaxation stress (RS), and transition temperature (TT).

sion used the same predictor preprocessing and response scaling as the ANN, with the regularisation parameter α selected from 0.01, 0.1, 1, 10, and 100 by minimising RMSE using fourfold inner cross-validation within each outer training dataset. Random forest regression used the same predictor preprocessing and was implemented with 200 trees, with all transformed predictor features considered at each split and a minimum terminal node size of one. R^2, MAE, and RMSE were calculated from the complete out-of-fold predictions in the same way as for the ANN. All model comparisons were therefore based on identical held-out treatment combinations in each outer fold.

To characterise variability in the ANN predictions arising from data partitioning and model refitting, the out-of-fold prediction for each treatment combination was retained from each of the 10 repeated outer cross-validation runs. Thus, each treatment combination had 10 out-of-fold predictions for each response. The mean and standard deviation (SD) of these predictions were then calculated for each treatment combination. The prediction standard deviation was used to characterise the variability of resampling-based model predictions and is shown as ± SD error bars in Fig. 7. Supplementary Data S1 provides the complete set of combination-level observed values, mean out-of-fold predictions, prediction standard deviations, and individual repeated out-of-fold prediction values. These SD values describe variability among repeated model fits and should not be interpreted as formal confidence intervals or external prediction intervals.

A more stringent grouped sensitivity analysis was also conducted to determine whether model performance depended on closely related mixture configurations appearing in both the training and test sets. Groups were defined by aggregate type, gradation, SBS modification, binder penetration grade, and binder content level. Thus, observations representing the same underlying mixture formulation but differing in specimen length or cooling rate were kept within the same fold. The same repeated nested cross-validation procedure was then applied with group-preserving folds.

Permutation importance reported in Table 16 was evaluated using only the held-out observations from the outer folds of the main repeated nested cross-validation. Each of the eight original predictors was randomly permuted 30 times, and predictor importance was quantified as the resulting decrease in test-set R^2. Importance values were aggregated across the 50 outer test folds and are reported as the mean ± standard deviation. These values quantify the model's dependence on each predictor in the dataset under investigation and should not be interpreted as evidence of physical or causal effects. Because correlated predictors may contain overlapping predictive information, their permutation importance values and rankings should be interpreted in the

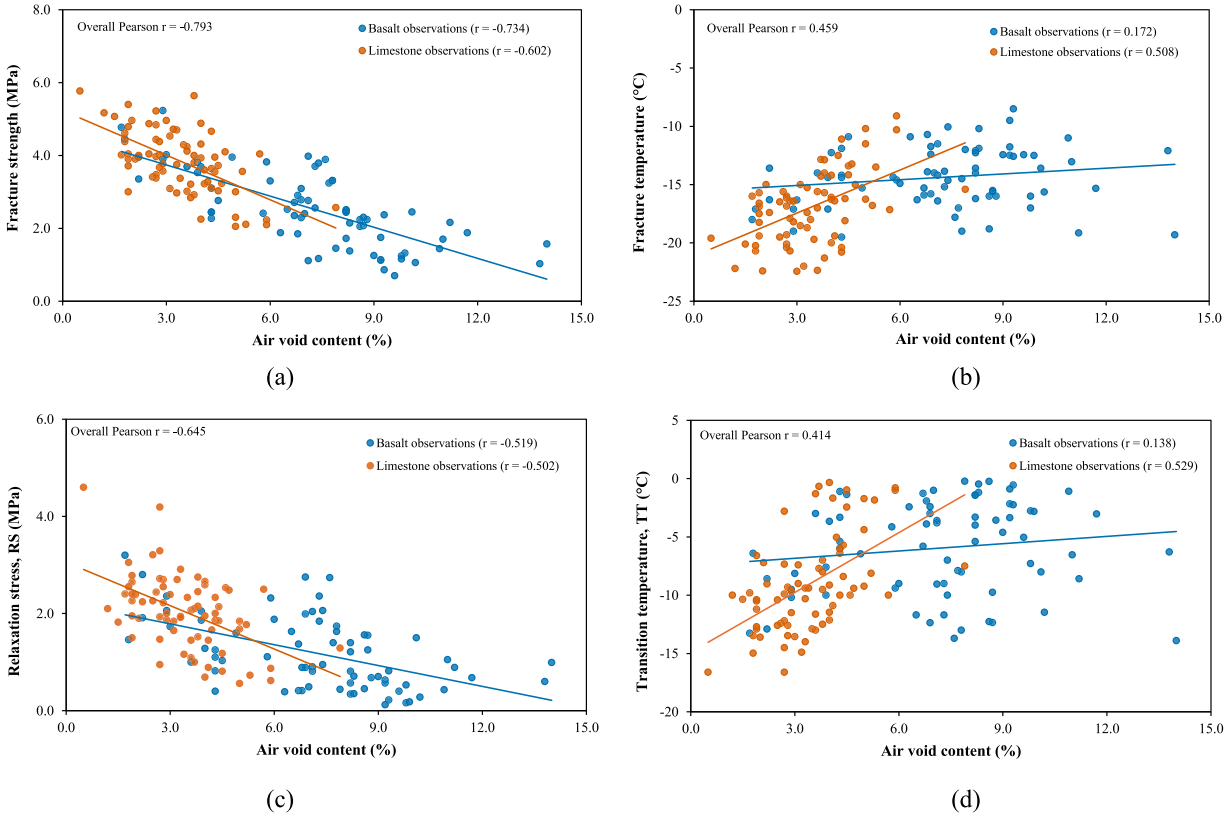

Fig. 5. Relationships between air void content and TSRST responses by aggregate type: (a) fracture strength, (b) fracture temperature, (c) relaxation stress, and (d) transition temperature.

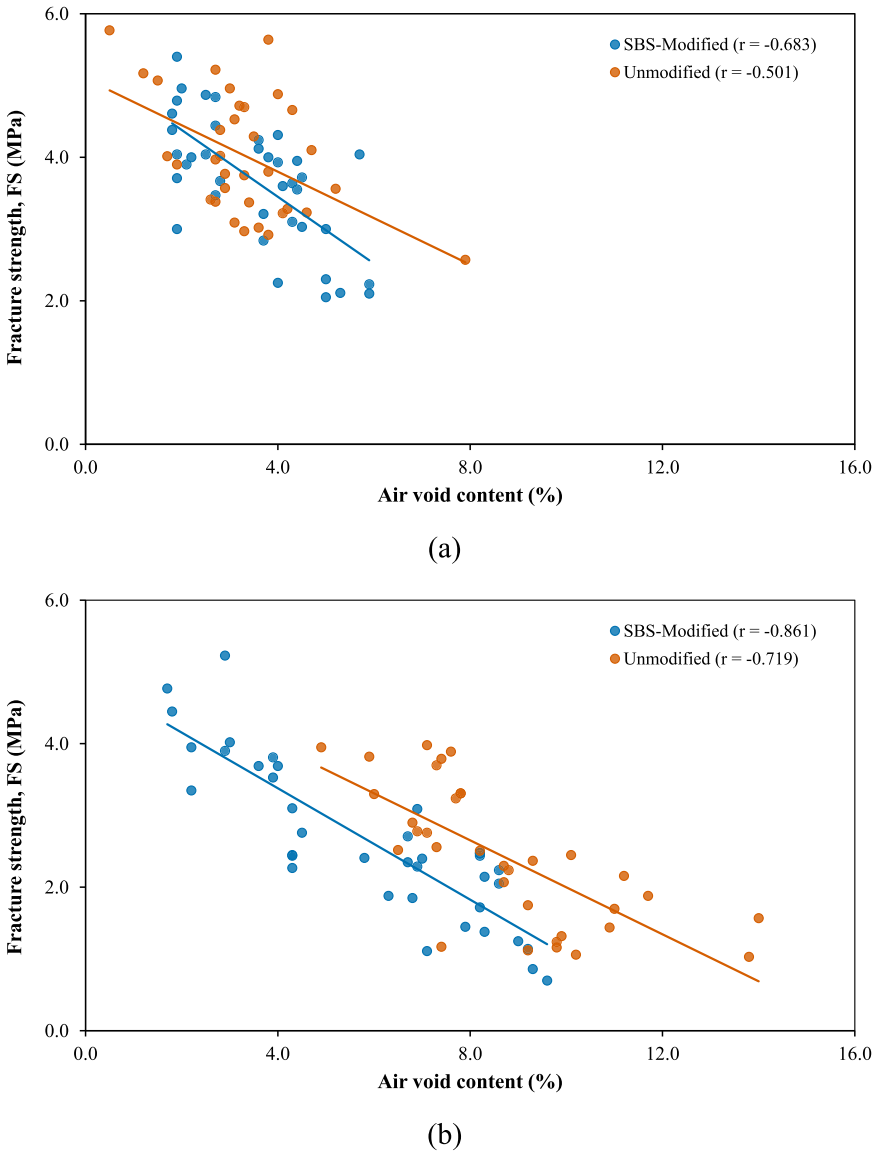

Fig. 6. Relationships between air void content and fracture strength for (a) limestone and (b) basalt mixtures, stratified by SBS modification.

context of the predictor structure of the current dataset. The ANN inputs, preprocessing procedures, and validation framework are summarised in Table 5.

4. Results

4.1. Descriptive statistics

The final analysed dataset comprised 144 specimen-level TSRST observations obtained from 72 compacted slabs representing 72 treatment combinations, with two beam specimens cut from each slab and tested separately. Descriptive statistics for the four TSRST responses are presented in Table 6. Fracture strength ranged from 0.70 MPa to 5.77 MPa, with a mean of 3.18 MPa and a standard deviation of 1.16 MPa. Fracture temperature ranged from −22.44 °C to −8.50 °C, with a mean of −15.65 °C and a standard deviation of 3.11 °C.

The mean relaxation stress was 1.61 MPa, with values ranging from 0.12 MPa to 4.60 MPa; the mean transition temperature was −7.50 °C, with values ranging from −16.60 °C to −0.23 °C. The observed dispersion reflects both variation associated with the investigated mixture design, specimen, and testing factors, and variation between replicate specimens. Therefore, the overall standard deviation should not be interpreted solely as a measure of TSRST repeatability.

4.2. Factorial ANOVA of TSRST responses

Separate main-effects ANOVA models were fitted for fracture strength (FS), fracture temperature (FT), relaxation stress (RS), and transition temperature (TT). Aggregate type, gradation, specimen length, SBS modification, binder penetration grade, binder content, and cooling rate were treated as categorical factors. Statistical significance was evaluated at $\alpha = 0.05$. Because the experimental programme followed a fractional factorial design, the analysis was focused on the main effects. The results should therefore be interpreted as factor-associated differences within the investigated design, recognising that unmodelled interaction effects may have contributed to the estimated main effects. The ANOVA results are summarised in Table 7. Aggregate type and binder content were significantly associated with all four TSRST response variables. Aggregate type was significant for FS, FT, RS, and TT (all $p < 0.001$) while binder content level was significant for FS, FT, and RS (all $p < 0.001$) and for TT ($p = 0.004$).

SBS modification was significantly associated with FS ($p = 0.019$) and FT ($p < 0.001$) but not with RS ($p = 0.294$) or TT ($p = 0.079$). Gradation was significant only for RS ($p = 0.031$); its associations with

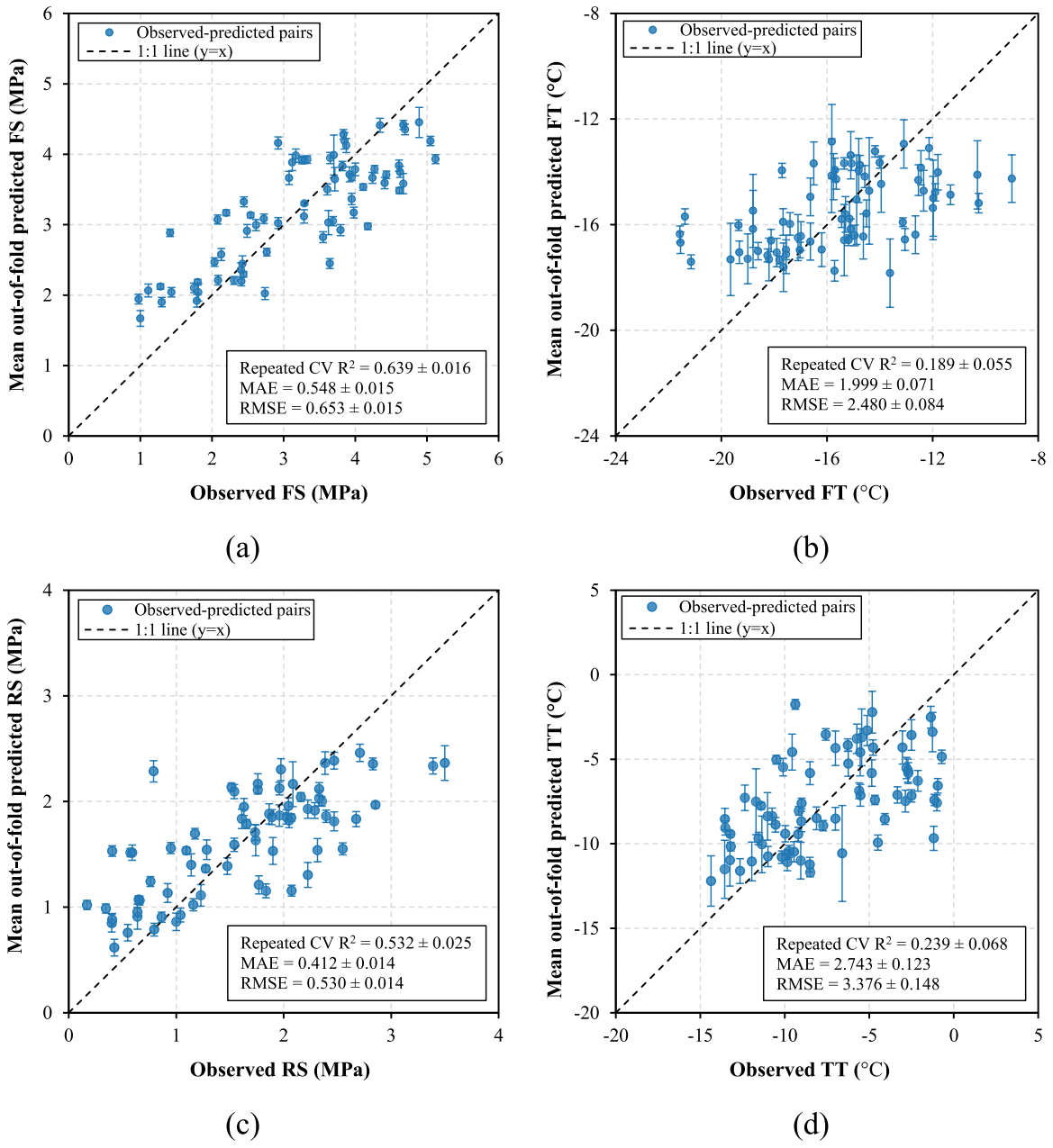

Fig. 7. Observed versus mean out-of-fold ANN predictions from 10 repetitions of nested fivefold cross-validation for: (a) fracture strength, (b) fracture temperature, (c) relaxation stress, and (d) transition temperature. Each point represents one treatment combination, and the predicted value is the mean of its 10 out-of-fold predictions. Vertical error bars represent ± SD across the 10 out-of-fold predictions for each treatment combination. Performance statistics shown in each panel are mean ± SD across the 10 complete cross-validation repetitions.

FS ($p = 0.061$), FT ($p = 0.325$), and TT ($p = 0.254$) were not statistically significant. Specimen length, binder penetration grade, and cooling rate were not significantly associated with any of the four TSRST responses.

4.3. *Effect of aggregate type*

Aggregate type was significantly associated with all four TSRST responses ($p < 0.001$; Table 7). The corresponding specimen-level means are presented in Table 8. Limestone specimens exhibited a mean fracture strength of 3.84 MPa, approximately 52% higher than that of basalt specimens (2.53 MPa) and fractured at a lower mean temperature (−16.92 °C versus −14.37 °C). Mean relaxation stress was also higher for limestone than for basalt (2.04 MPa versus 1.18 MPa) while mean transition temperature was lower (−9.05 °C versus −5.95 °C). These results indicate that the investigated limestone mixture system sustained higher thermal tensile stresses and reached lower characteristic temperatures than the basalt mixture system.

The aggregate groups also differed substantially in volumetric condition, with mean air void contents of 3.42% for the limestone specimens and 7.24% for the basalt specimens. As shown in Fig. 5(a), fracture strength decreased with increasing air void content in both the basalt ($r = -0.734$) and limestone ($r = -0.602$) groups, with an overall correlation of $r = -0.793$. Relaxation stress showed a similarly negative association, with correlations of $r = -0.519$ for basalt, $r = -0.502$ for limestone, and $r = -0.645$ for the complete dataset (see Fig. 5(c)).

Fracture and transition temperatures became less negative with

Table 5
Summary of ANN inputs, preprocessing, and validation framework.

Variable	Type	Levels or range	Encoding/model treatment
Aggregate type	Input	Limestone / Basalt	Categorical; one-hot encoded
Gradation	Input	Coarse / Fine	Categorical; one-hot encoded
Specimen length	Input	200 mm / 300 mm	Categorical; one-hot encoded
SBS modification	Input	Unmodified / SBS-modified	Categorical; one-hot encoded
Binder penetration grade	Input	50/70 / 70/100	Categorical; one-hot encoded
Binder content level	Input	OAC − 0.5% / OAC / OAC + 0.5%	Categorical; one-hot encoded
Cooling rate	Input	−5 / −10 / −20 °C/h	Numerical; standardised
Mean air void content	Input	1.1–13.9%	Continuous; standardised
Response	Output	FS / FT / RS / TT	Standardised within training folds
Candidate hidden layers	Model	–	(4), (8), or (8, 4)
Activation	Model	–	tanh
L2 regularisation, α	Model	–	0.01, 0.1, or 1.0
Optimiser	Model	–	Adam
Primary validation	Validation	–	10 × nested fivefold CV
Inner model selection	Validation	–	Fourfold CV; minimum RMSE
Grouped sensitivity analysis	Validation	–	Repeated nested grouped fivefold CV
Permutation importance	Interpretation	–	Held-out outer fold observations; 30 permutations
Prediction variability	Interpretation	–	Mean ± SD of 10 out-of-fold predictions for each treatment combination

Note: OAC = optimum asphalt content. All preprocessing operations were fitted using training data only.

Table 6
Descriptive statistics of the specimen-level TSRST responses.

Response	n	Mean	SD	Minimum	Maximum
Fracture strength, FS (MPa)	144	3.18	1.16	0.70	5.77
Fracture temperature, FT (°C)	144	−15.65	3.11	−22.44	−8.50
Relaxation stress, RS (MPa)	144	1.61	0.86	0.12	4.60
Transition temperature, TT (°C)	144	−7.50	4.36	−16.60	−0.23

Note: n = number of specimens; SD = standard deviation.

Table 7
ANOVA p-values for the TSRST response parameters.

Factor	FS	FT	RS	TT
Aggregate type	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
Gradation	$p = 0.061$	$p = 0.325$	$p = 0.031$	$p = 0.254$
Specimen length	$p = 0.071$	$p = 0.527$	$p = 0.960$	$p = 0.644$
SBS modification	$p = 0.019$	$p < 0.001$	$p = 0.294$	$p = 0.079$
Binder penetration grade	$p = 0.222$	$p = 0.140$	$p = 0.642$	$p = 0.453$
Binder content	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p = 0.004$
Cooling rate	$p = 0.078$	$p = 0.179$	$p = 0.608$	$p = 0.742$

Note: FS = fracture strength; FT = fracture temperature; RS = relaxation stress; TT = transition temperature. Statistical significance was evaluated at $\alpha = 0.05$.

Table 8
Specimen-level TSRST responses grouped by aggregate type. Values are mean ± SD.

Aggregate type	n	FS (MPa)	FT (°C)	RS (MPa)	TT (°C)	Air void content (%)
Limestone	72	3.84 ± 0.87	−16.92 ± 3.11	2.04 ± 0.76	−9.05 ± 4.15	3.42 ± 1.28
Basalt	72	2.53 ± 1.03	−14.37 ± 2.56	1.18 ± 0.73	−5.95 ± 4.02	7.24 ± 2.67

Note: n = number of specimens; FS = fracture strength; FT = fracture temperature; RS = relaxation stress; TT = transition temperature; SD = standard deviation.

increasing air void content (Fig. 5(b) and (d)), with overall correlations of $r = 0.459$ for FT and $r = 0.414$ for TT. The corresponding within-group correlations were $r = 0.508$ for FT and $r = 0.529$ for TT in the limestone group, compared with $r = 0.172$ and $r = 0.138$, respectively, in the basalt group. Previous studies have also identified aggregate type and air void content as important mixture-level factors affecting low-temperature fracture responses [9,20]. In TSRST testing, Velásquez et al. [35] reported that mixtures with 4% air voids were less susceptible to low-temperature cracking than those with 7% air voids. Accordingly, the aggregate-related ranking observed in the present study should not be interpreted as a general ranking of limestone and basalt; rather, it reflects the behaviour of two mixture systems that differed substantially in their volumetric conditions and, potentially, in aggregate packing and compactability.

The substantially higher air void content of the basalt specimens may have contributed to the observed differences in fracture strength, stress relaxation, fracture temperature, and transition temperature between the two aggregate types. However, air void content alone cannot establish the origin of these differences, and the present experimental programme was not designed to isolate the physicochemical or microstructural mechanisms responsible for aggregate-associated behaviour. Aggregate morphology, packing, and compactability, aggregate–binder interfacial characteristics, and the resulting internal void structure may all contribute to low-temperature behaviour at the mixture level. These characteristics were not quantified independently using microstructural or interfacial techniques in the present study. In addition, the experimental design did not include a comparison group involving interface modification or anti-stripping agents. Therefore, the contribution of aggregate–binder interfacial adhesion cannot be separated from differences in air void content, aggregate morphology, packing, compactability, or other characteristics of the mixture systems under investigation. A dedicated experiment that controls mixture volumetrics while varying interfacial modification would be needed to isolate the underlying mechanism. Consequently, the lower air void content and more favourable TSRST responses observed in the limestone mixtures cannot be attributed to any specific intrinsic property of limestone. The results should therefore be interpreted as reflecting differences between the investigated limestone and basalt mixture systems rather than as evidence of a distinct causal effect of aggregate mineralogy.

4.4. Effect of SBS modification

Table 9 presents the specimen-level TSRST responses grouped by binder modification. The factorial ANOVA showed that SBS modification was significantly associated with fracture strength ($p = 0.019$) and fracture temperature ($p < 0.001$), whereas its associations with relaxation stress ($p = 0.294$) and transition temperature ($p = 0.079$) were not statistically significant. The raw specimen-level mean fracture strengths of the SBS-modified and unmodified groups were similar, at 3.15 MPa and 3.22 MPa, respectively. However, the main-effects ANOVA identified a statistically significant SBS-associated difference after accounting

Table 9
Specimen-level TSRST responses and air void content grouped by binder modification. Values are mean ± SD.

Binder modification	n	FS (MPa)	FT (°C)	RS (MPa)	TT (°C)	Air void content (%)
Unmodified	68	3.22 ± 1.20	−16.31 ± 2.81	1.73 ± 0.87	−9.18 ± 3.67	6.00 ± 3.23
SBS-modified	76	3.15 ± 1.12	−15.05 ± 3.27	1.50 ± 0.85	−5.99 ± 4.40	4.73 ± 2.29

Note: n = number of specimens; FS = fracture strength; FT = fracture temperature; RS = relaxation stress; TT = transition temperature; SD = standard deviation.

for the other experimental factors in the model. SBS-modified specimens fractured at a mean temperature 1.26 °C higher than unmodified specimens (−15.05 °C versus −16.31 °C). Mean relaxation stress decreased from 1.73 MPa to 1.50 MPa while mean transition temperature shifted from −9.18 °C to −5.99 °C; however, neither response exhibited a statistically significant SBS-associated effect. Thus, SBS modification was significantly associated with FS and FT, but the observed mean responses did not indicate improved fracture strength or a shift towards lower fracture temperatures.

Previous studies have shown that the low-temperature performance of SBS-modified asphalt depends on the properties of the base binder, the SBS type and dosage, and other formulation variables, such as sulphur and processing oil contents [6,36]. Luo et al. [36] reported that SBS modification did not substantially improve the binder-level low-temperature properties whereas Hasan and Tarefder [37] found that mixtures prepared with binders containing 3% and 5% SBS exhibited better TSRST performance than mixtures prepared with the unmodified binder. Lin et al. [6] demonstrated that SBS type and content, sulphur content, and rubber-processing oil influence low-temperature performance and identified an optimum SBS content of 4.5% for the investigated mixture system.

Differences in air void content should also be considered when interpreting these results. The mean air void content was lower in the SBS-modified specimens than in the unmodified specimens (4.7% versus 6.0%). Because lower air void content would generally be expected to improve fracture resistance, the lower air void content of the SBS-modified specimens does not account for the absence of a consistent SBS-related improvement in the TSRST responses. Information on SBS dosage, polymer characteristics, and production procedures was unavailable because the SBS-modified binders were supplied ready to use. In addition, binder-level rheological and chemical characterisation, as well as aggregate–binder interfacial measurements, were not performed. The absence of a consistent SBS-related improvement in the present TSRST results may therefore reflect the specific combination of base–binder characteristics, SBS formulation and processing, mixture volumetrics, and other mixture-level factors. However, the available data cannot determine the relative contributions of these factors or establish the underlying micromechanisms. In particular, no conclusion regarding the formation or effectiveness of an SBS polymer network, or its influence on low-temperature interfacial behaviour, can be drawn from TSRST results alone. Consequently, the present SBS-associated findings apply only to the binder–mixture systems investigated and should not be generalised to SBS-modified asphalt mixtures in general.

4.5. *Aggregate type × SBS modification*

Table 10 and Fig. 6 present the TSRST responses to SBS modification separately for the limestone and basalt mixtures. For the limestone mixtures, the mean fracture strength was lower in the SBS-modified specimens than in the unmodified specimens (3.67 MPa versus 4.03 MPa), as was the mean relaxation stress (1.80 MPa versus 2.24 MPa). Mean fracture and transition temperatures also became less negative, shifting from −17.74 °C to −16.19 °C and from −10.85 °C to −7.43 °C, respectively. Thus, none of the four descriptive mean responses suggested improved low-temperature performance with SBS modification in the limestone mixture system.

For the basalt mixtures, the mean fracture strength was slightly higher in the SBS-modified specimens than in the unmodified specimens (2.63 MPa versus 2.42 MPa). However, the mean relaxation stress was slightly lower (1.15 MPa versus 1.22 MPa) while the mean fracture and transition temperatures were less negative in the SBS-modified specimens (−13.92 °C versus −14.72 °C and −4.55 °C versus −7.51 °C, respectively). Therefore, the apparent reversal in the direction of the SBS-associated response between the aggregate groups was restricted to fracture strength. Mean fracture temperature was less negative in the SBS-modified specimens for both aggregate types, consistent in direction with the significant main effect of SBS on FT. The subgroup differences in RS and TT remain descriptive because the corresponding SBS main effects were not statistically significant.

Differences in air void content provide important context for these subgroup responses. The difference in mean air void content between the modification groups was small for the limestone mixtures but substantially greater for the basalt mixtures. Mean air void content was similar in the unmodified and SBS-modified limestone specimens (3.30% and 3.53%, respectively). This pattern is also evident in Fig. 6: the limestone observations for the two modification groups largely overlap along the air void axis, whereas the unmodified basalt observations extend towards higher air void contents than the SBS-modified basalt observations. In contrast, the mean air void content was substantially lower in the SBS-modified basalt specimens than in the unmodified basalt specimens (5.94% versus 8.71%). A focused two-factor ANOVA identified a significant interaction between aggregate type and SBS modification for air void content ($F(1,140) = 23.35$, $p < 0.001$), indicating that the difference in air void content between SBS-modified and unmodified specimens varied by aggregate type.

As shown in Fig. 6, fracture strength exhibited a negative Pearson correlation with air void content in all four aggregate–modification subgroups. For limestone, the correlations were $r = -0.683$ for SBS-modified specimens and $r = -0.501$ for unmodified specimens. For basalt, the corresponding correlations were $r = -0.861$ and $r = -0.719$, respectively. Across most of the overlapping air-void range, the fitted regression line for the SBS-modified specimens lay below that for the unmodified specimens for both aggregate types. Thus, although the SBS-modified basalt specimens exhibited a slightly higher overall mean fracture strength, they also had substantially lower mean air void

Table 10
Specimen-level mean TSRST responses and air void content grouped by aggregate type and SBS modification.

Aggregate type	SBS modification	n	FS (MPa)	FT (°C)	RS (MPa)	TT (°C)	Air void content (%)
Limestone	Unmodified	34	4.03	−17.74	2.24	−10.85	3.30
Limestone	SBS-modified	38	3.67	−16.19	1.80	−7.43	3.53
Basalt	Unmodified	34	2.42	−14.72	1.22	−7.51	8.71
Basalt	SBS-modified	38	2.63	−13.92	1.15	−4.55	5.94

Note: n = number of specimens; FS = fracture strength; FT = fracture temperature; RS = relaxation stress; TT = transition temperature.

contents. The higher subgroup mean therefore cannot be interpreted as evidence of an independently beneficial effect of SBS modification. Overall, the aggregate-stratified TSRST responses remain descriptive, whereas the interaction between aggregate type and SBS modification for air void content was statistically significant. Previous studies have shown that low-temperature performance is influenced by SBS formulation characteristics, including polymer type and content, sulphur content, and processing oil [6] while aggregate type, binder grade and modification, binder content, and air void content can affect the low-temperature fracture resistance of asphalt mixtures [20].

4.6. Effect of binder content

Table 11 summarises the specimen-level TSRST responses and air void contents grouped by binder content level. The main-effects ANOVA showed that binder content level was significantly associated with all four TSRST responses. Mixtures prepared at OAC showed the highest mean FS and RS, at 3.41 MPa and 1.79 MPa, respectively. Their mean FT was −15.95 °C, which was only slightly more negative than that of the OAC + 0.5% group (−15.92 °C). The OAC − 0.5% group exhibited the lowest mean FS and RS and the least negative mean FT and TT. The mean air void content decreased from 6.87% at OAC − 0.5% to 5.13% at OAC and to 4.39% at OAC + 0.5%.

Because the investigated binder content levels were defined relative to the OAC of each base mixture, the findings demonstrate the importance of evaluating deviations from the mixture-specific OAC when interpreting low-temperature TSRST responses. Given the unequal numbers of specimens in the three binder content groups, the descriptive means should be interpreted alongside the main-effects ANOVA results, which consider the other experimental factors included in the model.

4.7. Effect of cooling rate

Table 12 summarises the specimen-level TSRST responses, grouped by cooling rate. The intermediate cooling rate of −10 °C/h exhibited the highest mean fracture strength and relaxation stress, at 3.40 MPa and 1.77 MPa, respectively, along with the lowest mean fracture and transition temperatures, at −16.11 °C and −7.82 °C. Mean FS and RS were lower at both −5 °C/h and −20 °C/h while mean FT and TT were less negative. These group means indicate a descriptive, non-monotonic response across the cooling rates investigated.

However, the ANOVA results showed that cooling rate was not significantly associated with fracture strength ($p = 0.078$), fracture temperature ($p = 0.179$), relaxation stress ($p = 0.608$), or transition temperature ($p = 0.742$). Although the result for fracture strength was the closest to the adopted significance threshold, it remained above $\alpha = 0.05$ and was therefore not considered statistically significant. Stress relaxation is an important component of the TSRST response [9]. From a viscoelastic perspective, a higher cooling rate generally reduces the time available for stress relaxation and thus accelerates the build-up of thermal stresses. Accordingly, an influence of cooling rate on the TSRST response is physically expected, although its magnitude depends on the material system and testing conditions [22,23]. For the mixture examined in their comparison of cooling rates, Pszczola et al. [22] found that increasing the cooling rate generally shifted fracture towards higher, less negative temperatures. The higher mean fracture temperature observed at −20 °C/h relative to −10 °C/h is consistent with this trend; nevertheless, the complete dataset did not exhibit a simple monotonic response. Moreover, the differences between the mean TSRST responses at the three cooling rates were small relative to the within-group variability. Cooling rate may therefore be considered a potentially relevant test parameter, although no statistically significant association with the TSRST responses was identified within the investigated −5 to −20 °C/h range and the present test configuration.

Table 11
Specimen-level TSRST responses and air void content grouped by binder content level. Values are mean ± SD.

Binder content level	n	FS (MPa)	FT (°C)	RS (MPa)	TT (°C)	Air void content (%)
OAC − 0.5%	38	2.70 ± 1.13	−14.85 ± 3.18	1.19 ± 0.71	−5.90 ± 4.10	6.87 ± 2.77
OAC	56	3.41 ± 1.06	−15.95 ± 2.78	1.79 ± 0.83	−7.91 ± 4.34	5.13 ± 2.75
OAC + 0.5%	50	3.30 ± 1.19	−15.92 ± 3.36	1.73 ± 0.91	−8.25 ± 4.35	4.39 ± 2.53

Note: OAC = optimum asphalt content; n = number of specimens; FS = fracture strength; FT = fracture temperature; RS = relaxation stress; TT = transition temperature; SD = standard deviation.

Table 12
Specimen-level TSRST responses grouped by cooling rate. Values are mean ± SD.

Cooling rate (°C/h)	n	FS (MPa)	FT (°C)	RS (MPa)	TT (°C)
−5	48	3.02 ± 1.16	−15.63 ± 3.41	1.52 ± 0.89	−7.41 ± 4.87
−10	48	3.40 ± 1.19	−16.11 ± 2.86	1.77 ± 0.91	−7.82 ± 4.08
−20	48	3.13 ± 1.11	−15.20 ± 3.04	1.54 ± 0.77	−7.26 ± 4.16

Note: n = number of specimens; FS = fracture strength; FT = fracture temperature; RS = relaxation stress; TT = transition temperature; SD = standard deviation.

4.8. Effects of gradation, specimen length, and binder penetration grade

The specimen-level descriptive results for gradation, specimen length, and binder penetration grade are shown in Table 13. Coarse-graded mixtures exhibited higher mean fracture strength and relaxation stress than fine-graded mixtures, at 3.49 MPa versus 2.87 MPa and 1.86 MPa versus 1.36 MPa, respectively. They also exhibited more negative mean fracture and transition temperatures, at −16.22 °C versus −15.08 °C and −8.59 °C versus −6.40 °C, respectively. However, only the association between gradation and relaxation stress was statistically significant ($p = 0.031$). The associations of gradation with fracture strength ($p = 0.061$), fracture temperature ($p = 0.325$), and transition temperature ($p = 0.254$) were not statistically significant. Therefore,

Table 13
Specimen-level TSRST responses grouped by gradation, specimen length, and binder penetration grade. Values are mean ± SD.

Factor	Level	n	FS (MPa)	FT (°C)	RS (MPa)	TT (°C)
Gradation	Coarse	72	3.49 ± 0.87	−16.22 ± 2.78	1.86 ± 0.75	−8.59 ± 4.21
	Fine	72	2.87 ± 1.32	−15.08 ± 3.33	1.36 ± 0.90	−6.40 ± 4.26
Specimen length	200 mm	70	3.27 ± 1.15	−15.79 ± 3.08	1.65 ± 0.92	−7.46 ± 4.71
	300 mm	74	3.11 ± 1.17	−15.51 ± 3.16	1.57 ± 0.81	−7.53 ± 4.03
Binder penetration grade	50/70	70	3.14 ± 1.12	−15.29 ± 3.28	1.63 ± 0.92	−7.01 ± 4.62
	70/100	74	3.23 ± 1.19	−15.99 ± 2.93	1.59 ± 0.80	−7.96 ± 4.08

Note: n = number of specimens; FS = fracture strength; FT = fracture temperature; RS = relaxation stress; TT = transition temperature; SD = standard deviation.

the higher FS and more negative FT and TT values observed for the coarse gradation should be interpreted as descriptive differences rather than evidence of statistically significant associations.

The differences between the two specimen lengths were minor. The 200 mm specimens had a slightly higher mean fracture strength than the 300 mm specimens, at 3.27 MPa and 3.11 MPa, respectively while their mean FT, RS, and TT values were similar. Specimen length was not significantly associated with FS ($p = 0.071$), FT ($p = 0.527$), RS ($p = 0.960$), or TT ($p = 0.644$). This result does not necessarily conflict with previous reports of specimen size effects in TSRST tests. In the present study, only the specimen length was varied from 200 to 300 mm while the cross-section was kept constant at 50 × 65 mm. In contrast, the specimen size comparison reported by Jung and Vinson [9] involved changes in both specimen length and cross-sectional dimensions, so the independent effect of length could not be isolated. Because specimen cross-section and heat-transfer conditions can influence thermal equilibration and temperature lag [24], maintaining the same cross-section may have reduced thermal differences between the two specimen lengths. The present result should therefore be interpreted as indicating the absence of a statistically detectable length-dependent difference within the investigated 200–300 mm range and test configuration rather than as evidence that specimen geometry is generally unimportant in TSRST.

Mixtures prepared with the 70/100 binder exhibited slightly lower mean fracture and transition temperatures than those prepared with the 50/70 binder, at −15.99 °C versus −15.29 °C and −7.96 °C versus −7.01 °C, respectively. Binder penetration grade was not significantly associated with FS ($p = 0.222$), FT ($p = 0.140$), RS ($p = 0.642$), or TT ($p = 0.453$). These differences should therefore be interpreted descriptively. Overall, of the three factors considered in this section, only the association between gradation and relaxation stress was statistically significant.

4.9. Cross-validated predictive performance and benchmark comparison

The predictive performance of the ANN models was evaluated using 10 repetitions of nested fivefold cross-validation, with hyperparameter selection performed independently within each outer training set. The resulting performance therefore reflects predictions for treatment combinations that were excluded from model fitting. Table 14 summarises the mean and standard deviation of R^2, MAE, and RMSE across the 10 full repetitions of cross-validation.

The FS model exhibited the strongest cross-validated performance, with a mean R^2 of 0.639 ± 0.016, an MAE of 0.548 ± 0.015 MPa, and an RMSE of 0.653 ± 0.015 MPa. The RS model showed moderate predictive performance, with an R^2 of 0.532 ± 0.025, an MAE of 0.412 ± 0.014 MPa, and an RMSE of 0.530 ± 0.014 MPa. In contrast, the temperature-based responses showed weaker cross-validated predictive performance. The FT model produced an R^2 of 0.189 ± 0.055, an MAE of 1.999 ± 0.071 °C, and an RMSE of 2.480 ± 0.084 °C. The TT model produced an R^2 of 0.239 ± 0.068, an MAE of 2.743 ± 0.123 °C, and an RMSE of 3.376 ± 0.148 °C. Thus, the cross-validated predictive performance was substantially stronger for FS and RS than for FT and TT.

The grouped sensitivity analysis produced similar results when complete underlying mixture formulations were excluded from the corresponding training folds. Mean grouped-cross-validation R^2 values were 0.613 ± 0.019 for FS, 0.152 ± 0.077 for FT, 0.499 ± 0.023 for RS, and 0.240 ± 0.073 for TT. Corresponding RMSE values were 0.676 ± 0.017 MPa, 2.534 ± 0.115 °C, 0.548 ± 0.013 MPa, and 3.373 ± 0.156 °C, respectively. The relatively small differences between the primary and grouped validation results indicate that the estimated performance did not depend strongly on closely related variants of specimen length or cooling rate of the same mixture formulation appearing in both the training and test data. Nevertheless, both validation procedures were conducted entirely within the material types and factor levels used in the experimental programme and therefore do not demonstrate extrapolation beyond the investigated domain. Accordingly, these performance statistics reflect internal out-of-fold validation rather than independently and externally validated predictive performance.

Benchmark comparisons using identical, repeated outer cross-validation folds are presented in Table 15. The ANN did not consistently outperform the benchmark models. For FS, ridge regression and random forest regression both yielded higher mean R^2 values (0.706) than the ANN (0.639), with lower RMSE values of 0.588 and 0.589 MPa, respectively, compared with 0.653 MPa for the ANN. For FT, the random

Table 15
Cross-validated comparison of ANN and benchmark models using identical repeated outer folds.

Response	Model	R^2	MAE	RMSE
FS	ANN	0.639 ± 0.016	0.548 ± 0.015 MPa	0.653 ± 0.015 MPa
	Ridge regression	0.706 ± 0.032	0.502 ± 0.026 MPa	0.588 ± 0.030 MPa
	Random forest	0.706 ± 0.020	0.479 ± 0.023 MPa	0.589 ± 0.020 MPa
FT	ANN	0.189 ± 0.055	1.999 ± 0.071 °C	2.480 ± 0.084 °C
	Ridge regression	0.241 ± 0.045	1.910 ± 0.053 °C	2.399 ± 0.070 °C
	Random forest	0.317 ± 0.069	1.859 ± 0.100 °C	2.274 ± 0.116 °C
RS	ANN	0.532 ± 0.025	0.412 ± 0.014 MPa	0.530 ± 0.014 MPa
	Ridge regression	0.511 ± 0.027	0.435 ± 0.016 MPa	0.541 ± 0.015 MPa
	Random forest	0.540 ± 0.037	0.415 ± 0.019 MPa	0.525 ± 0.021 MPa
TT	ANN	0.239 ± 0.068	2.743 ± 0.123 °C	3.376 ± 0.148 °C
	Ridge regression	0.339 ± 0.052	2.650 ± 0.114 °C	3.145 ± 0.123 °C
	Random forest	0.536 ± 0.026	2.143 ± 0.054 °C	2.638 ± 0.074 °C

Note: Values are mean ± SD across 10 complete outer cross-validation repetitions. All models were evaluated using identical outer folds. Ridge regularisation was selected using fourfold inner cross-validation within each outer training set. Random forest regression was used as a predefined tree-based benchmark with 200 trees. FS = fracture strength; FT = fracture temperature; RS = relaxation stress; TT = transition temperature.

Table 14
Cross-validated performance of the ANN models.

Response	Repeated nested CV R^2	MAE	RMSE	Grouped CV R^2	Grouped MAE	Grouped RMSE
FS	0.639 ± 0.016	0.548 ± 0.015 MPa	0.653 ± 0.015 MPa	0.613 ± 0.019	0.572 ± 0.015 MPa	0.676 ± 0.017 MPa
FT	0.189 ± 0.055	1.999 ± 0.071 °C	2.480 ± 0.084 °C	0.152 ± 0.077	2.055 ± 0.083 °C	2.534 ± 0.115 °C
RS	0.532 ± 0.025	0.412 ± 0.014 MPa	0.530 ± 0.014 MPa	0.499 ± 0.023	0.427 ± 0.012 MPa	0.548 ± 0.013 MPa
TT	0.239 ± 0.068	2.743 ± 0.123 °C	3.376 ± 0.148 °C	0.240 ± 0.073	2.736 ± 0.154 °C	3.373 ± 0.156 °C

Note: Values are mean ± SD across 10 complete outer cross-validation repetitions. These SD values describe variation in the overall performance metrics among repeated data partitions and are distinct from the treatment-combination-level prediction SDs shown in Fig. 7 and Supplementary Data S1.

forest achieved the best performance ($R^2 = 0.317$; RMSE = 2.274 °C), compared with $R^2 = 0.189$ and RMSE = 2.480 °C for the ANN. For RS, ANN and random forest showed similar performance, with R^2 values of 0.532 and 0.540 and RMSE values of 0.530 and 0.525 MPa, respectively. The largest difference occurred for TT, for which the random forest achieved $R^2 = 0.536$ and RMSE = 2.638 °C, compared with $R^2 = 0.239$ and RMSE = 3.376 °C for the ANN. These results indicate that relative predictive performance depended on the response, and that the ANN did not provide a consistent predictive advantage over the benchmark models.

Fig. 7 compares the observed responses with the mean out-of-fold predictions for each treatment combination across the 10 repeated nested cross-validation runs. Vertical error bars represent ± SD of the 10 out-of-fold predictions for each combination and therefore show the variability arising from repeated data partitioning and model refitting. To characterise this resampling-based prediction variability, the standard deviation of the 10 out-of-fold predictions was calculated separately for each treatment combination. Across the 72 combinations, the mean combination-specific predictive standard deviation was 0.081 MPa for FS, 0.676 °C for FT, 0.074 MPa for RS and 0.712 °C for TT. The corresponding ranges were 0.024–0.284 MPa, 0.152–2.017 °C, 0.030–0.213 MPa, and 0.172–2.839 °C, respectively. These SD values quantify the variation arising from repeated data partitioning and model refitting within each response and should not be interpreted as confidence intervals or compared directly across responses that use different units or scales. Taken together with the lower cross-validated R^2 values, the FT and TT results indicate more limited predictive performance for the temperature-based responses. Complete out-of-fold prediction results at the treatment combination level are provided in Supplementary Data S1.

4.10. Held-out permutation feature importance analysis

Table 16 presents the permutation importance values obtained using only the held-out observations from the outer cross-validation folds. Predictor importance varied substantially across the responses. Mean air void content was the highest-ranked predictor for all four responses, producing mean reductions in held-out R^2 of 0.863 for FS, 0.364 for FT, 0.625 for RS, and 0.307 for TT. This consistent ranking indicates that ANN predictions, evaluated in held-out outer folds, depended strongly on the volumetric condition characterised by the mean air void content.

Table 16
Held-out permutation importance of the predictors for the four ANN models.

Predictor	FS	FT	RS	TT
Mean air void content	0.863 ± 0.165 (1)	0.364 ± 0.153 (1)	0.625 ± 0.149 (1)	0.307 ± 0.153 (1)
Aggregate type	0.051 ± 0.039 (2)	0.026 ± 0.054 (3)	0.060 ± 0.045 (2)	0.021 ± 0.044 (4)
Binder content level	0.028 ± 0.044 (3)	−0.003 ± 0.056 (6)	0.055 ± 0.056 (3)	0.007 ± 0.056 (5)
SBS modification	0.025 ± 0.021 (4)	0.017 ± 0.048 (4)	0.050 ± 0.037 (4)	0.135 ± 0.090 (2)
Gradation	0.016 ± 0.039 (5)	−0.014 ± 0.087 (7)	0.040 ± 0.053 (5)	0.098 ± 0.127 (3)
Specimen length	−0.000 ± 0.010 (6)	−0.021 ± 0.071 (8)	0.003 ± 0.015 (7)	−0.012 ± 0.096 (8)
Binder penetration grade	−0.014 ± 0.027 (7)	0.007 ± 0.037 (5)	0.007 ± 0.039 (6)	−0.008 ± 0.031 (7)
Cooling rate	−0.028 ± 0.029 (8)	0.027 ± 0.071 (2)	−0.027 ± 0.029 (8)	0.005 ± 0.082 (6)

Note: Values are reported as mean reduction in held-out R^2 ± SD across the 50 outer test folds. Predictor ranks are shown in parentheses. Negative or near-zero values indicate no stable held-out predictive contribution and should not be interpreted as causal effects.

For FS and RS, aggregate type ranked second, with mean R^2 reductions of 0.051 and 0.060, respectively; and binder content ranked third. For TT, SBS modification and gradation ranked second and third, respectively. For FT, the importance of predictors other than mean air void content was relatively low and unstable, consistent with the limited overall cross-validated performance of the FT model.

The permutation importance rankings did not fully correspond to the ANOVA main effect results because the two analyses address different aspects of the data. The main-effects ANOVA estimates mean differences associated with each factor while accounting for the other experimental factors included in the model. Permutation importance instead measures the reduction in predictive performance when a predictor is perturbed in held-out data.

Several lower-ranked predictors had mean permutation importance values that were close to or below zero. These values indicate that randomly permuting these variables did not consistently reduce predictive performance on the held-out dataset. They should therefore be interpreted as evidence of limited or unstable predictive contributions within the fitted ANN models rather than as beneficial or inverse physical effects. More generally, permutation importance quantifies how strongly a model depends on its features and should not be interpreted as evidence of causal relationships. The rankings should also be interpreted with caution because several predictors in the current experimental dataset are correlated. In particular, mean air void content varies systematically across certain materials and mixture groups, so predictive information may be shared between air void content and other ANN inputs. Permuting a correlated predictor can therefore change the apparent importance attributed to it and to related predictors. The reported rankings therefore describe model dependence within the current predictor structure and should not be interpreted as a hierarchy of physical importance.

5. Discussion

Aggregate type and binder content level were the only factors significantly associated with all four TSRST responses. Limestone mixtures exhibited higher FS and RS and more negative FT and TT than basalt mixtures, indicating higher stress levels and lower characteristic temperatures under the conditions investigated. However, the limestone specimens also had substantially lower mean air void content. The observed differences should therefore be interpreted as differences between the investigated limestone and basalt mixture systems rather than as effects of aggregate mineralogy alone. Therefore, the present results do not demonstrate a general low-temperature performance advantage of limestone over basalt. Accordingly, aggregate selection and control of mixture volumetrics should be considered together in low-temperature mixture design.

Binder content level was significantly associated with all four TSRST responses, although the magnitude and direction of the associated differences varied among the response parameters. Mixtures prepared at OAC exhibited the highest mean FS and the lowest mean FT. This pattern suggests that departures from the mixture-specific optimum may influence low-temperature fracture response through associated changes in effective binder content, asphalt film thickness, and mixture volumetrics [20,30] and may also affect thermal stress development [7,31]. Because the investigated binder content levels were defined relative to the OAC of each base mixture, the findings demonstrate the importance of evaluating deviations from the mixture-specific OAC when interpreting low-temperature TSRST responses. Gradation was significantly associated only with RS, whereas neither specimen length nor binder penetration grade was significantly associated with any of the four TSRST responses.

SBS modification was significantly associated with FS and FT but not with RS or TT, and the observed responses did not indicate a consistent improvement in low-temperature performance. The overall mean FS was similar for SBS-modified and unmodified mixtures, whereas the SBS-

modified mixtures fractured at a warmer, less negative mean temperature. However, the slightly higher mean FS observed for the SBS-modified basalt subgroup coincided with a substantial reduction in the mean air void content. The response to SBS modification should therefore be interpreted in the context of the resulting combined aggregate–binder–volumetric condition. Because information on SBS dosage and production conditions was unavailable and binder rheology and interfacial adhesion were not measured, the underlying mechanisms cannot be conclusively established from the present data. Accordingly, these observations are specific to the SBS-modified binder–mixture systems investigated and should not be interpreted as evidence that SBS modification generally offers no low-temperature benefit.

The mean TSRST responses varied non-monotonically with cooling rate. The −10 °C/h condition showed the highest mean FS and RS and the most negative mean FT and TT, although cooling rate was not significantly associated with any of the four responses. The observed pattern should therefore be interpreted descriptively rather than as evidence of an optimal cooling rate. The non-monotonic cooling rate response should not be interpreted as evidence that the cooling rate is physically unimportant. Previous studies have shown that thermal stress development can depend on the cooling rate [22,23] and that chamber configuration, airflow, and specimen geometry can influence temperature equilibration during TSRST [24]. This dependence is consistent with the viscoelastic nature of asphalt mixtures, because faster cooling generally allows less time for stress relaxation before further thermal contraction. In the present study, however, cooling was not controlled solely by the chamber air set-point; temperature feedback from four specimen-mounted RTDs was used to regulate the prescribed cooling programme, and the resulting cooling rate was monitored during testing. Thus, the lack of statistical significance observed across −5, −10, and −20 °C/h reflects the response of the mixtures investigated and the test configuration rather than the nominal chamber set points alone. The cooling system had also been checked using an internally instrumented dummy specimen before testing, further reducing the likelihood that the prescribed rates reflected only the chamber air temperature.

A limitation of the present analysis is that RS and TT depend on the procedure used to identify the transition to the approximately linear part of the stress–temperature response. Although the same automated smoothing and R^2-based terminal segment selection procedure was applied consistently to all specimens, a dedicated robustness assessment using alternative smoothing or segment selection settings was not performed. Consequently, some contribution of the processing procedure to the absolute RS and TT values cannot be ruled out. These responses should therefore be interpreted primarily as consistently derived comparative indicators within the current experimental programme rather than as material properties independent of the processing conditions.

The absence of a significant association with specimen length should likewise be interpreted in the context of the geometry studied. Unlike earlier TSRST size comparisons, in which both specimen length and cross-sectional dimensions varied [9], the present study varied only specimen length, maintaining a constant 50 × 65 mm cross-section. Since specimen cross-section and heat-transfer conditions can influence thermal equilibration [24], the two lengths may therefore have experienced broadly similar thermal conditions under the present test configuration. Nevertheless, modest length-related differences cannot be ruled out, and targeted experiments that vary specimen dimensions independently would provide a stronger assessment of geometrical scaling effects.

The fractional factorial design also limits causal interpretation of the observed factor-associated differences. Although the statistical models identify significant associations between selected factors and TSRST responses, these associations should not be interpreted as definitive evidence that any individual factor independently causes the observed responses. Unmeasured interactions among mixture, specimen, and testing variables may contribute to the observed differences. Follow-up experiments specifically designed to distinguish main effects from interaction effects would be required to support stronger causal or mechanistic inferences.

The factorial ANOVA was conducted at the specimen level using 144 individual TSRST measurements from 72 compacted slabs. For each treatment combination, the two beam specimens were cut from the same slab and therefore were not independent mixture-production or slab-level experimental replicates. Nevertheless, the beams were physically distinct specimens, had individually measured volumetric properties, and were tested separately in the TSRST. Retaining their individual responses preserves the observed within-slab specimen variability; however, the specimen-level ANOVA does not explicitly model the dependence arising from their shared slab of origin. Accordingly, the reported statistical significance should be interpreted with caution and in conjunction with the 72 distinct treatment combinations in the experimental design.

The repeated nested cross-validation provided an internal assessment of ANN performance based on held-out treatment combinations. The FS model exhibited the strongest cross-validated predictive performance ($R^2 = 0.639 \pm 0.016$), followed by RS ($R^2 = 0.532 \pm 0.025$). In contrast, the FT and TT models explained considerably smaller proportions of the variation in withheld observations, with R^2 values of 0.189 ± 0.055 and 0.239 ± 0.068, respectively. The ANN framework therefore showed moderate cross-validated predictive performance for the stress-based responses FS and RS, but limited performance for the temperature-based responses FT and TT, given the available predictors and data. The resampling-based prediction standard deviations further characterised the variability in the out-of-fold predictions for each response. Because the four responses had different units and scales, these standard deviation values were not directly compared across the response types. The grouped validation provided an additional robustness check by ensuring that observations from the same underlying mixture formulation did not appear in the training and test folds simultaneously. The grouped R^2 values of 0.613, 0.152, 0.499, and 0.240 for FS, FT, RS, and TT, respectively, were close to those obtained in the ordinary repeated nested cross-validation. This indicates that the primary performance estimates were not driven solely by closely related variants in cooling rate or specimen length that appeared in both the training and test subsets. However, all withheld formulations remained within the aggregate types, binder systems, gradations, binder content levels, and testing ranges represented in the experimental programme. The validation therefore assesses internal generalisation to withheld combinations or formulations within the investigated domain and does not establish extrapolation to new asphalt materials or external engineering conditions.

The benchmark comparison further showed that the ANN did not offer a consistent predictive advantage over the alternative models. Ridge regression and random forest both outperformed the ANN for FS, while random forest achieved the best performance for FT and TT. For RS, the ANN and random forest produced very similar cross-validated performance. Accordingly, relative predictive performance depended on the response, and the ANN results should be interpreted as a complementary modelling approach rather than as evidence that ANNs are inherently superior to alternative statistical or tree-based models. Held-out permutation analysis identified mean air void content as the highest-ranking predictor in all four ANN models. This finding is consistent with experimentally observed relationships between volumetric condition and TSRST response; however, the permutation importance values should be interpreted as measures of model dependence rather than physical or causal significance. Because the mean air void content is associated with several material and mixture characteristics in the present dataset, shared predictive information among the correlated inputs may affect both the magnitude and ranking of their permutation importance. The remaining predictors generally showed much smaller

and less stable held-out importance values, particularly for FT and TT. Differences between the ANOVA and ANN importance rankings are not contradictory, because ANOVA evaluates adjusted factor-associated mean differences, whereas permutation importance quantifies the predictive model's dependence on each input in held-out data.

The predictive models evaluated through internal cross-validation may therefore provide preliminary screening information within the experimental domain being investigated. The benchmark results indicate that the ANN should not be regarded as consistently superior to simpler alternative models, and its performance should instead be assessed on a case-by-case basis. Independent external validation using separately prepared mixtures, additional material sources, or factor levels not represented during model fitting is required before wider engineering application.

The experimental programme was designed primarily to quantify macroscopic TSRST responses across a broad set of mixture design, specimen, and cooling variables rather than to identify the underlying microstructural or physicochemical mechanisms. Accordingly, aggregate morphology and surface characteristics, internal pore morphology and connectivity, aggregate–binder interfacial properties, binder chemical composition and low-temperature rheology, and SBS morphology and distribution were not characterised separately. Techniques such as X-ray computed tomography for internal void structure, microscopic examination of fractures and aggregate–binder interfaces, surface energy measurements, and rheological or chemical characterisation of the binders could provide complementary evidence to distinguish the mechanisms responsible for the observed aggregate and SBS-related responses. Without such measurements, the present results establish statistically supported mixture-level associations but should not be interpreted as direct evidence for specific microscopic mechanisms. Future work should include direct interfacial characterisation and controlled comparison groups, with and without interface modification at comparable air void contents, to distinguish adhesion-related effects from volumetric and aggregate-structure effects.

The mixtures were subjected to three hours of oven conditioning for short-term ageing before slab compaction; however, no additional long-term laboratory ageing treatment was applied before TSRST testing. Consequently, the present results reflect the short-term ageing of the laboratory mixtures but do not capture the oxidative ageing that accumulates during long-term pavement service. Long-term ageing may alter binder stiffness, stress relaxation, and fracture response, so the magnitude and relative importance of the factors identified here may differ in materials that have aged in service. Future work should evaluate TSRST responses after controlled long-term ageing to determine the ageing sensitivity of these mixtures.

6. Conclusions

This study presents a structured comparison of seven factors related to material, mixture design, specimen configuration and testing within a common TSRST experimental framework, using 144 specimens representing 72 treatment combinations and evaluating both fracture and pre-fracture responses. Because the study employed a fractional factorial experimental design, the reported findings should be interpreted as factor–response associations within the investigated test matrix rather than as independent causal effects. The main conclusions are as follows:

1. Only aggregate type and binder content level were significantly associated with all four TSRST responses. Limestone mixtures exhibited higher mean FS and RS values, together with more negative mean FT and TT values, than basalt mixtures. However, because the limestone specimens also had substantially lower air void contents, the observed differences should be interpreted as differences between the investigated mixture systems rather than as effects of aggregate mineralogy alone.
2. Binder content level was significantly associated with all four TSRST responses. Mixtures prepared at their mixture-specific OAC exhibited the highest mean FS and RS, while their mean FT was only slightly more negative than that of the OAC + 0.5% mixtures. These findings demonstrate the importance of evaluating deviations from the mixture-specific OAC when assessing low-temperature TSRST behaviour.
3. SBS modification was significantly associated with FS and FT but not with RS or TT, and the observed mean responses did not indicate a consistent improvement in low-temperature performance. The aggregate-stratified comparisons were descriptive and the modest increase in mean FS for SBS-modified basalt mixtures was accompanied by substantially lower mean air void content than in the corresponding unmodified mixtures. The present data do not permit these aggregate- or SBS-associated responses to be attributed to specific microstructural, interfacial, or binder-level mechanisms.
4. Within the investigated −5 to −20 °C/h range and given test configuration, cooling rate was not significantly associated with any of the four TSRST outcomes. Although the −10 °C/h group exhibited the highest mean FS and RS and the most negative mean FT and TT, the pattern across the three cooling rates was non-monotonic and should not be interpreted as evidence for an optimal cooling rate. Gradation was significantly associated only with RS, whereas specimen length was not significantly associated with any of the four TSRST responses within the investigated 200–300 mm range with constant cross-section configuration. Binder penetration grade was not significantly associated with any of the four TSRST response variables.
5. Repeated nested cross-validation yielded mean R^2 values of 0.639, 0.189, 0.532, and 0.239 for FS, FT, RS, and TT, respectively. A grouped validation, in which entire mixture formulations were withheld, yielded similar R^2 values of 0.613, 0.152, 0.499, and 0.240. Thus, the ANN framework showed moderate cross-validated predictive performance for FS and RS, and limited performance for FT and TT. Benchmark comparisons using identical outer folds showed that the ANN did not consistently outperform the benchmark models. Ridge regression and random forest performed better for FS, random forest provided the strongest predictive performance for FT and TT, and ANN and random forest showed comparable performance for RS. Held-out permutation analysis identified mean air void content as the highest-ranked predictor in all four ANN models; however, these rankings describe model dependence within the correlated experimental dataset and should not be interpreted as indicating physical or causal significance. The models should be regarded as preliminary screening tools within the experimental domain investigated and should not be extrapolated to new materials or factor levels without independent external validation.

These conclusions apply to the laboratory mixtures subjected to short-term ageing; their applicability to long-term field-aged pavement materials requires further validation under controlled long-term ageing conditions.

CRediT authorship contribution statement

Adnan Qadir: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Murat Guler:** Writing – review & editing, Validation, Supervision, Conceptualization. **Yalcin Karakaya:** Writing – review & editing, Writing – original draft, Validation, Investigation, Data curation, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT for language editing and grammar improvement. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Funding

This work was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under project no. 105M193.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors gratefully acknowledge the laboratory facilities and technical staff of the Department of Civil Engineering at Middle East Technical University for their assistance with specimen preparation, testing and data acquisition. The authors also acknowledge the Turkish General Directorate of Highways for its valuable support during the experimental phase of this study.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.conbuildmat.2026.148751.

Data Availability

Data will be made available on request.

References

[1] W. Shen, D.J. Kirkner, Thermal cracking of viscoelastic asphalt-concrete pavement, J. Eng. Mech. 127 (7) (2001) 700–709, https://doi.org/10.1061/(ASCE)0733-9399(2001)127:7(700).
[2] Vinson, T.S., Janoo, V.C., & Haas, R.C.G. (1989). Summary report: Low temperature and thermal fatigue cracking (Report No. SHRP-A/IR-90-001). Strategic Highway Research Program, National Research Council.
[3] N. Tapsoba, H. Baaj, C. Sauzéat, H. Di Benedetto, M. Ech, 3D analysis and modelling of thermal stress restrained specimen test (TSRST) on asphalt mixes with RAP and roofing shingles, Constr. Build. Mater. 120 (2016) 393–402, https://doi.org/10.1016/j.conbuildmat.2016.05.092.
[4] Z. Sun, J. Yi, Y. Li, Y. Tan, S. Li, S. Yang, Y. Zhang, Z. Yue, Evaluating the low-temperature performance of asphalt mixtures based on the contraction and relaxation properties, Constr. Build. Mater. 412 (2024) 134874, https://doi.org/10.1016/j.conbuildmat.2024.134874.
[5] J. Tang, Y. Fu, T. Ma, B. Zheng, Y. Zhang, X. Huang, Investigation on low-temperature cracking characteristics of asphalt mixtures: A virtual thermal stress restrained specimen test approach, Constr. Build. Mater. 347 (2022) 128541, https://doi.org/10.1016/j.conbuildmat.2022.128541.
[6] P. Lin, W. Huang, Y. Li, N. Tang, F. Xiao, Investigation of influence factors on low temperature properties of SBS modified asphalt, Constr. Build. Mater. 154 (2017) 609–622, https://doi.org/10.1016/j.conbuildmat.2017.06.118.
[7] D. Rys, M. Jaczewski, M. Pszczola, P. Jaskula, W. Bankowski, Effect of bitumen characteristics obtained according to EN and Superpave specifications on asphalt mixture performance in low-temperature laboratory tests, Constr. Build. Mater. 231 (2020) 117156, https://doi.org/10.1016/j.conbuildmat.2019.117156.
[8] Y. Luo, H. Wu, W. Song, J. Yin, Y. Zhan, J. Yu, S.A. Wada, Thermal fatigue and cracking behaviors of asphalt mixtures under different temperature variations, Constr. Build. Mater. 369 (2023) 130623, https://doi.org/10.1016/j.conbuildmat.2023.130623.
[9] D. Jung, T.S. Vinson, Thermal stress restrained specimen test to evaluate low-temperature cracking of asphalt-aggregate mixtures, Transp. Res. Rec. 1417 (1993) 12–20.
[10] O. Shafaghat, P. Hajikarimi, F.M. Nejad, F.F. Tehrani, Heterogeneous Finite Element Modeling of TSRST to Analyze Key Parameters Affecting the Low-Temperature Viscoelastic Behavior of HMA, Results Eng. 30 (2026) 110805, https://doi.org/10.1016/j.rineng.2026.110805.
[11] S. Rahman, A. Bhasin, A. Smit, Exploring the use of machine learning to predict metrics related to asphalt mixture performance, Constr. Build. Mater. 295 (2021) 123585, https://doi.org/10.1016/j.conbuildmat.2021.123585.
[12] S. Ghafari, M. Ehsani, F.M. Nejad, Prediction of low-temperature fracture resistance curves of unmodified and crumb rubber modified hot mix asphalt mixtures using a machine learning approach, Constr. Build. Mater. 314 (2022) 125332, https://doi.org/10.1016/j.conbuildmat.2021.125332.
[13] A. Qadir, Investigation of low temperature cracking in asphalt concrete pavement (Doctoral dissertation). Middle East Technical University, ProQuest Dissertations & Theses Global, Ankara, Turkey., 2010.
[14] Y. Karakaya, A. Qadir, A. Arabzadeh, M. Guler, Effects of mixture design variables on low-temperature cracking and thermal fatigue resistance of asphalt mixtures, Constr. Build. Mater. 542 (2026) 148004, https://doi.org/10.1016/j.conbuildmat.2026.148004.
[15] H.A. Tabatabaee, R. Velasquez, H.U. Bahia, Modeling thermal stress in asphalt mixtures undergoing glass transition and physical hardening, Transp. Res. Rec. 2296 (1) (2012) 106–114, https://doi.org/10.3141/2296-11.
[16] D. Wang, A.C. Falchetto, C. Riccardi, M.P. Wistuba, Investigation on the low temperature properties of asphalt binder: glass transition temperature and modulus shift factor, Constr. Build. Mater. 245 (2020) 118351, https://doi.org/10.1016/j.conbuildmat.2020.118351.
[17] Lytton, R.L., Uzan, J., Fernando, E.G., Roque, R., Hiltunen, D., & Stoffels, S.M. (1993). Development and Validation of Performance Prediction Models and Specifications for Asphalt Binders and Paving Mixes, SHRP-A-357, Strategic Highway Research Program, National Research Council, Washington, DC.
[18] J.T. Christison, D.W. Murry, K.O. Anderson, Stress prediction and low-temperature fracture susceptibility of asphaltic concrete pavements, Proc. Assoc. Asph. Paving Technol. 41 (1972) 494–523.
[19] M. Zeng, H.U. Bahia, H. Zhai, M.R. Anderson, P. Turner, Rheological modeling of modified asphalt binders and mixtures (with discussion), J. Assoc. Asph. Paving Technol. 70 (2001) 403–441.
[20] X. Li, M.O. Marasteanu, A. Kvasnak, J. Bausano, R.C. Williams, B. Worel, Factors study in low-temperature fracture resistance of asphalt concrete, J. Mater. Civil. Eng. 22 (2) (2010) 145–152, https://doi.org/10.1061/(ASCE)0899-1561(2010)22:2(145).
[21] A.C. Falchetto, K.H. Moon, M.P. Wistuba, Development of a simple correlation between bending beam rheometer and thermal stress restrained specimen test low-temperature properties based on a simplified size effect approach, Road. Mater. Pavement Des. 18 (sup2) (2017) 339–351, https://doi.org/10.1080/14680629.2017.1305147.
[22] M. Pszczola, C. Szydlowski, M. Jaczewski, Influence of cooling rate and additives on low-temperature properties of asphalt mixtures in the TSRST, Constr. Build. Mater. 204 (2019) 399–409, https://doi.org/10.1016/j.conbuildmat.2019.01.148.
[23] A.K. Apeagyei, E.V. Dave, W.G. Buttlar, Effect of cooling rate on thermal cracking of asphalt concrete pavements, Asph. Paving Technol. -Proc. 77 (2008) 709–738.
[24] D. Steiner, B. Hofko, On the influence of basic thermodynamics on thermal cracking resistance of asphalt mixtures in cooling tests, RILEM Tech. Lett. 3 (2018) 1–7, https://doi.org/10.21809/rilemtechlett.2018.54.
[25] A.F. Braham, W.G. Buttlar, M.O. Marasteanu, Effect of binder type, aggregate, and mixture composition on fracture energy of hot-mix asphalt in cold climates, Transp. Res. Rec. 2001 (1) (2007) 102–109, https://doi.org/10.3141/2001-12.
[26] M.M. Aliha, H. Behbahani, H. Fazaeli, M.H. Rezaifar, Study of characteristic specification on mixed mode fracture toughness of asphalt mixtures, Constr. Build. Mater. 54 (2014) 623–635, https://doi.org/10.1016/j.conbuildmat.2013.12.097.
[27] J. Du, D. Ren, C. Ai, J. Zhang, Y. Qiu, Effect of aggregate gradation on crack propagation in asphalt mixtures at low temperatures based on the Eshelby equivalent inclusion theory, Constr. Build. Mater. 290 (2021) 123181, https://doi.org/10.1016/j.conbuildmat.2021.123181.
[28] F. Xiao, X. Xiao, J. Li, Low-temperature fracture behavior of railway asphalt concretes under semi-circular bending: Experimental and numerical investigation, Constr. Build. Mater. 402 (2023) 132990, https://doi.org/10.1016/j.conbuildmat.2023.132990.
[29] F. Dong, W. Zhao, Y. Zhang, J. Wei, W. Fan, Y. Yu, Z. Wang, Influence of SBS and asphalt on SBS dispersion and the performance of modified asphalt, Constr. Build. Mater. 62 (2014) 1–7, https://doi.org/10.1016/j.conbuildmat.2014.03.018.
[30] M. Oshone, D. Ghosh, E.V. Dave, J.S. Daniel, J.M. Voels, S. Dai, Effect of mix design variables on thermal cracking performance parameters of asphalt mixtures, Transp. Res. Rec. 2672 (28) (2018) 471–480, https://doi.org/10.1177/0361198118797826.
[31] M. Akentuna, S.S. Kim, M. Nazzal, A.R. Abbas, M.S. Arefin, Study of the thermal stress development of asphalt mixtures using the Asphalt Concrete Cracking Device (ACCD), Constr. Build. Mater. 114 (2016) 416–422, https://doi.org/10.1016/j.conbuildmat.2016.03.207.
[32] J. Leukel, L. Scheurer, V. Sugumaran, Machine learning models for predicting physical properties in asphalt road construction: A systematic review, Constr.

Build. Mater. 440 (2024) 137397, https://doi.org/10.1016/j.conbuildmat.2024.137397.
[33] D. Mirzaiyanrajeh, E.V. Dave, J.E. Sias, P. Ramsey, Developing a prediction model for low-temperature fracture energy of asphalt mixtures using machine learning approach, Int. J. Pavement Eng. 24 (2) (2023) 2024185, https://doi.org/10.1080/10298436.2021.2024185.
[34] M. Atakan, K. Yıldız, Prediction of Marshall design parameters of asphalt mixtures via machine learning algorithms based on literature data, Road. Mater. Pavement Des. 25 (3) (2024) 454–473, https://doi.org/10.1080/14680629.2023.2213774.
[35] R. Velásquez, M. Turos, M. Marasteanu, N. Gibson, T. Clyne, Investigation of low-temperature cracking in asphalt mixtures using TSRST. Pavement Cracking: Mechanisms, Modeling, Detection, Testing and Case Histories, CRC Press, 2008, pp. 405–414.
[36] Z.G. Luo, S.Y. Wang, B. Diao, Exploration on low-temperature properties of SBS modified asphalt for pavement, CICTP 2012 Multimodal Transp. Syst. Conv. safe Cost. -Eff. Effic. (2012) 3112–3121, https://doi.org/10.1061/9780784412442.317.
[37] M.A. Hasan, R.A. Tarefder, Laboratory Investigation of Low-Temperature Performance of the SBS Modified Mixtures. International Airfield and Highway Pavements Conference 2019, American Society of Civil Engineers, Reston, VA, 2019, pp. 109–113, https://doi.org/10.1061/9780784482469.011.